Dirk Pijpops* and Dirk Speelman

Alternating argument constructions of Dutch psychological verbs: A theory-driven corpus investigation

Abstract: This article presents the results of a corpus study of the Dutch psychological verbs *ergeren* 'to annoy', *interesseren* 'to interest', *storen* 'to disturb', and *verbazen* 'to amaze'. These verbs exhibit a syntactic alternation between their seemingly synonymous transitive and reflexive argument constructions, as in *Elizabeth ergert John* vs. *John ergert zich aan Elizabeth* (both: 'Elizabeth annoys John'). On the basis of current theoretical insights, four hypotheses are formulated predicting the language user's preferred argument construction. It is argued that the popular agentivity hypothesis, as proposed in studies by, for instance, Dowty, Langacker, and Zaenen, should be broken up into the token- and type-level agentivity hypotheses. Both agentivity hypotheses come with different theoretical entailments, and make distinct predictions about the quantitative data. These data confirm the token-level agentivity hypothesis, while not doing the same for the type-level agentivity hypothesis. Additionally, it is found that stimuli and experiencers that are heavier in terms of informational weight both prompt the use of the reflexive construction, and that the individual preferences of the verbs could not be predicted based on their historical semantic development.

1 Introduction

Psychological verbs, or psych verbs, are a group of verbs denoting a mental state or attitude, which show a wide range of both inter- and intralingual variation in

*Corresponding author: Dirk Pijpops, Research Foundation Flanders (FWO), Department of Linguistics, KU Leuven, PO Box 3308, Blijde-Inkomststraat 21, 3000 Leuven, Belgium,
E-mail: dirk.pijpops@kuleuven.be
Dirk Speelman, Department of Linguistics, KU Leuven, PO Box 3308, Blijde-Inkomststraat 21, 3000 Leuven, Belgium, E-mail: dirk.speelman@kuleuven.be

argument realization (Croft 1993; Verhoeven 2010).[1] Examples of such syntactic variation in English are the alternations between *fear* and *frighten* in (1) and *like* and *please* in (2). The participant who causes the mental state – *Elizabeth* in (1) and (2) – is usually called the *stimulus*, whereas the participant who experiences the mental state – *John* – is named the *experiencer* (see Dowty 1991; Croft 1993; Klein and Kutscher 2005).[2] In this way, following Dowty (1991: 579), an initial distinction can be made between constructions which have the experiencer as their subject (experiencer-subject) and those which have the stimulus as their subject (stimulus-subject).

(1)   Experiencer-subject: *John fears Elizabeth.*
        Stimulus-subject: *Elizabeth frightens John.*

(2)   Experiencer-subject: *John likes Elizabeth.*
        Stimulus-subject: *Elizabeth pleases John.*

This distinction, however, is too coarse-grained to do justice to the wide range of argument constructions in which these verbs appear cross-linguistically (Croft 1993; Van de Velde 2004: 55–58).[3] The rich syntactic variation poses a problem for linking theories of argument realization, i. e., theories which aim to link a verb's semantic representation to its formal argument realization, such as approaches linking thematic with θ-roles (Baker 1997), Dowty's Argument Selection Principles (Dowty 1991), the mappings of the Macroroles Actor and Undergoer in Role and Reference Grammar (Van Valin 2004), or the Lexical Mapping Theory of Lexical Functional Grammar (Bresnan and Kanerva 1989; Bresnan 2001). Broadly speaking, this problem can be summarized as follows: if a verb's argument realization can be predicted on the basis of its semantics, then why do psychological verbs that express a very similar mental state, such as

---

1 The psychological verbs are also known as *mental verbs* (Croft 1986, 1993), *psychological predicates* (Postal 1970), *flip verbs* (Rogers 1974), and *experience predicates* (Van de Velde 2014).
2 We will continue using the terms *stimulus* and *experiencer* for practical reasons and to remain in line with previous work on the psychological verbs. However, we do not mean to attribute any theoretical status to them (e. g., as thematic roles or similar).
3 For a treatment of the psychological verbs in Croatian, see Saravanja (2011), in Czech Filip (1996), in Present-day English Grafmiller (2013), in Old English de la Cruz (1994), in Present-day Dutch Bennis (1986, 2004) and Zaenen (1993), in Middle Dutch van der Horst (1985) and Van de Velde (2004), in French Herschensohn (1992), Bouchard (1995) and Kelling (2003), in German Primus (2004) and Klein and Kutscher (2005), in Greek Kordini (2001), in Icelandic Barðdal (2009), in Italian Belletti and Rizzi (1988), in Korean Kang (1986) and Yang (1994), in Mandarin Cheung and Larson (2015), and in Romanian Alexiadou and Iordăchioaia (2014).

English *fear–frighten* or *like–please*, exhibit a different – even opposite – argument realization?[4]

Because any theory of argument realization pursuing full coverage of the facts is confronted with this question at some point, the treatment of psychological verbs has attracted much theoretical attention (Belletti and Rizzi 1988; Grimshaw 1990; Bouchard 1995; Pesetsky 1995; Vanhoe 2002; Primus 2004; Klein and Kutscher 2005). As noted by Grafmiller (2013: 262–269), however, this work has primarily relied on introspective judgments, be it occasionally coupled with typological evidence (Croft 1993). This has resulted in an imbalance between theoretical attention and quantitative research, with the former being predominant. Quantitative work has been especially lacking when compared to another central domain of variation in argument realization, namely the dative alternation (e. g., Bresnan et al. 2007; Colleman 2009; Theijssen 2012; Szmrecsanyi et al. 2016). It is this relative imbalance in research into the psychological verbs that we hope to redress, by supplying quantitative data and linking it to theoretical studies. Such data may then provide a sound empirical base on which to further develop theory formation. In doing so, we follow earlier work by Van de Velde (2004), Levin and Grafmiller (2012), and Grafmiller (2013), among others.

One of the main ways of dealing with the problem of the psychological verbs is through some form of the agentivity hypothesis (e. g., in Grimshaw 1990; Dowty 1991; Langacker 1991; Croft 1993; Zaenen 1993; Pesetsky 1995; Vanhoe 2002, see Section 2). The precise implementation of the hypothesis obviously differs for each theoretical framework, but in general, it involves the following principle: since the agentivity of the participants in a mental state – stimulus or experiencer – can vary, variation arises as to which one is assigned subject position. The most agentive of the two participants in the mental state to be expressed, then, is realized as the subject. However, we will claim that it is important to distinguish two levels at which this hypothesis may operate.

The first is provisionally called the *type level*. At this level, the meaning element of agentivity is part of the lexical meaning of the verb. More specifically, the type-level agentivity hypothesis claims that a verb whose lexical meaning entails either a more agentive stimulus or a more agentive experiencer will be more likely to realize said participant in subject position. It is thus the verb, or rather, the lexical meaning of the verb, that selects its argument construction.

---

4 This is an intralingual formulation of the problem. If linking theories claim to have crosslingual validity – and they often do (Hopper and Thompson 1980; Baker 1997; Van Valin 2004) – the problem is amplified by crosslingual variation in the argument realization of the psych verbs, e. g., English *like* vs. German *gefallen* and Spanish *gustar* (Dowty 1991; Klein and Kutscher 2005; Vanhoe 2002).

The second level is called the *token level*. At this level, agentivity is not part of the lexical meaning of the verb, but rather belongs to the meaning of the argument construction. As such, it is not the verb that selects its argument construction, but the language user. More specifically, the token-level agentivity hypothesis holds that if two argument constructions are possible to express a particular utterance, language users will put that participant in subject position that they wish to portray as more agentive. Agentive meaning is thus added separately to the compositional meaning of the whole utterance by the argument construction employed, and does not follow directly from the lexical meaning of the verb.

The alternation examined in the present study is one that allows us to easily discriminate between these levels (see Sections 3.1 and 3.2). In a number of psychological verbs in Dutch, the experiencer can be encoded as either subject or direct object. One such verb is *ergeren* 'to annoy', which may appear in either of two argument constructions: one is called the reflexive construction, as in (3), and the other is called the transitive construction, as in (4).[5] In the reflexive construction, the experiencer takes up subject position and the stimulus is realized as prepositional object, while in the transitive construction, the stimulus occupies subject position and the experiencer serves as direct object. The same alternation can also be found in German and French.[6]

(3)   Reflexive construction (experiencer-subject):[7]
      John ergert   zich   aan Elizabeth.
      John annoys himself to   Elizabeth
      'Elizabeth annoys John.'

---

5 This alternation is not an idiosyncratic characteristic of a few verbs, but a more general phenomenon occurring with many psychological verbs in Dutch. The following verbs show examples of the reflexive as well as the transitive construction, which can readily be found on the internet: *Amuseren* 'to amuse', *bedroeven* 'to sadden', *benieuwen* 'to make curious', *berouwen* 'to regret', *ergeren* 'to annoy', *frustreren* 'to frustrate', *generen* 'to embarrass', *interesseren* 'to interest', *irriteren* 'to irritate', *ontroeren* 'to move emotionally', *opwinden* 'to arouse', *plezieren* 'to make happy', *spijten* 'to regret', *storen* 'to disturb', *verbazen* 'to amaze', *verbijsteren* 'to baffle', *verblijden* 'to gladden', *verdrieten* 'to grieve', *vergenoegen* 'to content', *verheugen* 'to rejoice', *vermaken* 'to entertain', *verontwaardigen* 'fill with indignation', *vervelen* 'to bore', *verwonderen* 'to surprise'.
6 For example, with German *ärgern* 'to annoy' and French *irriter* 'to irritate'.
7 Not all may agree that this argument construction should be considered truly reflexive, as the reflexive pronoun is a fixed part of the construction, and cannot be replaced by a non-reflexive element. Doing so would in fact turn the construction into a transitive one, with a transitive meaning, as in (4). Still, the label "reflexive construction" clearly identifies the construction in (3) as distinct from (4).

(4)   Transitive construction (stimulus-subject):
       Elizabeth ergert   John.
       Elizabeth annoys John
       'Elizabeth annoys John.'

To native speakers of Dutch, both constructions seem interchangeable; there is no salient difference in meaning or use. In what follows, we will attempt to discover what makes Dutch language users opt for (3) or for the alternative in (4). To do so, we will not look for a single difference in meaning or use between the two constructions, but rather take a multivariate approach, trying to discern statistical patterns in the use of the reflexive and transitive constructions of various psychological verbs. By revealing gradual differences in meaning or use, such patterns could then tell us how and for what purposes language users utilize this difference in argument structure. The present study thus subscribes to the tradition of alternation research, as, for instance, in Bresnan et al. (2007), Speelman and Geeraerts (2009), Levin and Grafmiller (2012), and Pijpops and Van de Velde (2014). This means we will start from the assumption and null-hypothesis that there are in fact no such significant patterns to be found, and that language users use both constructions interchangeably.

    This article is structured as follows. Section 2 provides a short overview of some of the solutions offered by a theory of argument realization in solving the problem of the commonly occurring syntactic variation in psychological verbs. Here, we will also further outline the difference between the type- and token-level agentivity hypothesis. This will allow us to infer a number of hypotheses concerning the behavior of the alternation under scrutiny, in Section 3. Next, Section 4 describes the extraction and selection of the data, as well as the addition of several variables to the dataset. Section 5 presents the analysis of the dataset making use of logistic regression. The results of this analysis are interpreted in Section 6. The final section, then, contains the conclusions.

2 Theoretical background

Section 2.1 sketches the various approaches that linguists have taken to deal with the syntactic variation in psychological verbs, and it indicates how the present study relates to the other solutions. This survey is, of course, far too concise to do justice to all the work that has already been done on this topic, and has no intention of being exhaustive. It merely aims to make explicit some of the assumptions underlying the current investigation. For a more comprehensive overview of

the literature, we refer to Levin and Rappaport Hovav (2005) and Grafmiller (2013). In Section 2.2, the distinction between the type- and token-level agentivity hypothesis is further developed. While this distinction is in principle theory-neutral, we will employ terminology from construction grammar to describe it.

2.1 Solutions to the psych verbs problem

There are two main ways to deal with the psych verbs problem, while maintaining some form of linking theory of argument realization.[8] The first way is to agree that there is no difference in meaning between the various argument realizations of the psych verbs, or at least that this difference in meaning cannot be adequately accounted for in the linking system employed. The argument variation is then attributed to some other cause.

A first option is to search for this "other cause" in formal syntax. For the Dutch psych verbs, den Besten (1982), Bennis (1986: 134–147, 2004), and Hoekstra (1987) argue that a single syntactic deep structure for the various realizations of the verbs has to be assumed, yet that this structure does not assign an external role (see Bennis 1986: 138). As a result, movement operations are needed, which result in variation at the surface level. Belletti and Rizzi (1988, 291–295) do posit differing deep structures, but their analysis has met resistance from Grimshaw (1990), Bouchard (1995), Iwata (1995), Pesetsky (1995), Arad (1996), and Chung (1999).

A second option to account for the variation in argument realization without making reference to a difference in meaning is simply to specify the choice of argument structure lexically. This is what has been proposed by Vanhoe (2002: 383) for, e. g., the Spanish verbs *gustar* 'to please' and *temer* 'to fear', and by Van Valin (2004: 76–77) for, e. g., the German verbs *mögen* 'like' and *gefallen* 'please', the latter of which is specified to be *macrorole-intransitive*. These lexical specifications are often only used in conjunction with a linking system in order to account for a number of exceptional cases. The linking system then handles the majority of the psych verb variation in terms of meaning differences, as is done by Vanhoe's proto-agent and -patient properties (Vanhoe 2002: 379) and by Van Valin's Default Macrorole Assignment Principles (Van Valin 2004).

Finally, a third option is to claim that the variation is caused by a difference in meaning, but that this difference goes beyond the normal linking system. This is what has been proposed by Grimshaw (1990: 22–25), who attributes the same

---

8 Not all researchers agree that the psych verbs present a problem at all. For one, Bouchard (1995) argues against a special treatment of the psych verbs.

θ-roles to experiencer-subject and experiencer-object verbs, but ascribes the different argument realizations to a difference in prominence between the participants.

The second way of dealing with the psych verbs problem is to claim that, in fact, a meaning difference is detected by the linking system, causing it to select varying argument realizations. This is where some form of the agentivity hypothesis often comes into play.[9] In Pesetsky (1995) and Iwata (1995), the agentivity hypothesis is embodied through a difference between the role *Causer*, which corresponds to the stimulus when it is realized as subject, and the role called *Emotive Object* (Iwata 1995: 99–100), corresponding to the stimulus when realized as object. In the works of Lakoff, Hopper and Thompson, and Langacker, notions such as agentivity (Lakoff 1977), transitivity (Hopper and Thompson 1980), and subject, direct object, etc. (Langacker 1991) are attributed prototypical status. In these accounts, the argument realization of a verb is determined by its correspondence to these prototypes, with the more agentive participant more likely to take up subject position. According to Croft (1991: 216–225, 1993; see also Croft 1998, 2012), an important distinction lies in the causal structure of the verbs. For causative mental verbs, like *to please*, the starting point of the causal chain would be the stimulus, whereas for stative mental verbs, such as *to like*, it could be either the stimulus or the experiencer, resulting in variation.

A central proposal in the work on argument realization is the work by Dowty (1991), which has found acceptance in several theoretical frameworks, such as θ-theory (Baker 1997), Lexical Functional Grammar (Zaenen 1993; Ackerman and Moore 2001; Vanhoe 2002; Kelling 2003), Optimality Theory (Primus 2002, 2004, 2006), and Construction Grammar (Goldberg 1995: 116–117). According to Dowty (1991), the argument position for each participant is determined by the relative degree to which each participant corresponds to a proto-agent or proto-patient. The more properties of this proto-agent and proto-patient a participant collects, the more likely it is to take up subject or object position, respectively. As mentioned by Dowty (1991: 600), these properties correspond to the factors that are typically mentioned in the literature, like volitional involvement, sentience, causality, movement, incremental theme, etc. Important in this respect is that language users can vary the weight of each property to suit their communicative needs.

---

9 This is not always the case, however. For instance, in Arad's (1996) proposal, the varying argument realizations are due to aspectual projections, with the basic event type of, e. g., *fear* being atelic and the basic event type of *frighten* being telic. However, due to the many so-called *outside modifications* that are possible, this proposal is hard to operationalize. For a more in-depth discussion on the problems of Arad's proposal, see Rappaport Hovav and Levin (2002: 277–278) and Grafmiller (2013: 17–21, 29–31, 40–73).

So far, we have used agentivity and the agentivity hypothesis as an umbrella term, subsuming causality, proto-roles, Langacker's *flow of energy* (Langacker 1991), and several other proposals that have not been mentioned, such as Jackendoff's action tier (Jackendoff 1990: 139–142). In what follows, we will continue to do so. However, this is certainly not meant to imply that the differences between these accounts are negligible or irrelevant. Rather, we employ the umbrella term *agentivity hypothesis* because we believe the data presented here are not well-suited to allow differentiation between these proposals, and because it is not our goal to prove or refute a single account. As mentioned above, our goal is rather to improve the quantitative foundation of this theoretical work, and to check whether, and how, agentivity influences the argument realization in Dutch psych verbs. In the next section, we will therefore differentiate between the two forms of the agentivity hypothesis that do make different predictions regarding our data. In the remainder of this paper, we will adopt Grafmiller's (2013: 219) definition of agentivity as a working definition (italics ours): *"When a situation does involve an agent, that event is said to instantiate the property of 'agentivity'. Agentivity in this sense is a property of events (or states); an agentive event involves the willful control of the event by some participant."* For a more elaborate characterization of agentivity and a discussion of earlier definitions, see Grafmiller (2013: 211–261).

A critical discussion of some of solutions offered as a first way to come to grips with the psych verbs problem, can be found in Grafmiller (2013: 10–73). The set-up of the present alternation study implies that we follow the second route: we will search for differences in use and meaning between varying argument realizations of psych verbs, and claim that these differences are the very raison d'être of the variation. That is, we claim that this variation exists not because of some formal syntactic property or movement, but rather because it is useful for language users to convey these differences when expressing mental states or events (Croft 2000; Bybee 2010). Such differences do not necessarily need to be differences in meaning, although these have been the main focus of most theoretical studies; it is also possible that they turn out to be differences in register (Plevoets 2008), lectal value (Colleman 2010; Geeraerts et al. 2010), or information structure (Goldberg 2006: 221). At the very least, such differences outside the domain of semantics should not be excluded a priori (see Section 4.2).

These usage-based views are compatible with the framework of construction grammar (Diessel 2015). As such, we will use the terminology of this framework in the remainder of the paper, more specifically the terminology of the variety of construction grammar put forward in Goldberg (1995, 2006), as it is focused on argument realization.

Finally, it is worth noting that the present study does not directly intend to resolve the psych verbs problem outlined above. Rather, it will attempt to apply the reasoning followed in the second way of dealing with the psych verbs problem, in order to make predictions regarding the alternation at issue. This is what is done in Section 2.2, and especially in 3.1 and 3.2.

2.2 Type and token level

In Section 1, the distinction was introduced between the type and token level at which the agentivity hypothesis may operate. The type-level agentivity hypothesis is concerned with the lexical meaning of a verb, that is, the meaning that is shared by all its occurrences. If this lexical meaning implies an agent-like experiencer, and a patient-like stimulus, the verb is likely to exhibit an experiencer-subject argument realization, and vice versa. Most linking mechanisms, such as θ-theory (Baker 1997) or the Macrorole Assignment Principles (Van Valin 2004), are primarily to be situated at this level. For instance, in Dowty's (1991: 579–581) work, which builds on Croft (1986), the verb *to please* has a stimulus-subject realization because the lexical meaning of *to please* entails one proto-agent and one proto-patient property for the experiencer, and only one proto-agent property and no proto-patient properties for the stimulus. The stimulus therefore has the strongest claim to subjecthood.

In the construction grammar view put forward in, for instance, Goldberg (1995, 2006), argument realization is handled through the compatibility between the lexical construction of the verb – carrying its lexical meaning – and the argument structure construction (henceforth called argument construction). The argument construction may specify the conditions that the lexical meaning of the verb needs to satisfy in order to be compatible with it (van Trijp 2015: 621–624). To investigate these conditions, verb disposition may be studied, i. e., the preferences of verbs to appear with a particular argument construction (Stallings et al. 1998). In this way, it can be shown that the ditransitive construction is more compatible with verbs whose lexical meaning expresses a "caused reception", like *to give*, while the prepositional dative construction is more compatible with verbs expressing a "caused motion", like *to bring* (Goldberg 1992; Gries and Stefanowitsch 2004; Colleman 2009). When this reasoning is applied to the alternation under scrutiny, it could be expected that verbs whose lexical meaning implies a more agentive experiencer, like *interesseren* 'to interest', will be more compatible with the reflexive construction, whereas verbs whose lexical meaning implies a less agentive experiencer, like *verbazen* 'to amaze', will be more compatible with the transitive construction (see Section 3.1).

It should be noted, however, that this compatibility account alone does not necessarily entail that argument constructions independently contribute to the compositional meaning of the whole utterance. First and foremost, this meaning element of agentivity is part of the lexical meaning of the verb.

Unlike the type-level agentivity hypothesis, the token-level agentivity hypothesis relates to the concrete meaning that a language user wishes to express in a particular context. Suppose a language user wanted to express a particular mental event or state – and would already have a verb selected for this purpose – but he/she also wanted to stress that the stimulus was exceptionally agentive. To convey this information, he/she would place the stimulus in subject position, i. e., employ a stimulus-subject argument construction. Note that in this case, it is the language user who determines the argument realization, not the verb. In addition, the meaning element of agentivity is not part of the lexical meaning of the verb, but is contributed independently by the argument construction to the compositional meaning of the utterance.[10]

Following this line of reasoning for the alternation in psych verbs, it could be predicted that, if a stimulus in a particular utterance is (construed as being) more agentive than in another, it would have a higher chance of appearing in subject position, i. e., the language user would more likely have opted for a stimulus-subject argument construction, such as the transitive construction.

Both the theoretical literature on the agentivity hypothesis and the quantitative studies available often do not strictly discriminate between the type and token levels. Rather, the influence of agentivity at the token level is seen as the natural outcome of its influence at the type level. For instance, the various linking mechanisms are designed to link a verb to its argument construction based on the lexical meaning of the verb, i. e., they are designed to operate at the type level, but they are often, in turn, used to license the agentivity hypothesis at the token level (Dowty 1991: 579–581; Levin and Grafmiller 2012: 21–24). Furthermore, in line with Goldberg (1995, 2006: 220–221); the very reason why some argument constructions are more compatible with certain verbs, i. e., the type level, is precisely because these argument constructions carry meaning independently of these verbs, which language users use to compositionally build the meaning of their whole utterances, i. e., the token level. It is also difficult to disentangle the two levels for alternations in which a different

---

10 It is emphasized in construction grammar studies focusing on the token level that verbs and argument constructions may freely combine with one another (Goldberg 2006: 220–221). Even if the verb and an argument construction are in principle not semantically compatible, they may be coerced, should the language user find it useful to do so (van Trijp 2015: 613–615).

argument construction can only be obtained by switching the verb, such as English *like–please* or *fear–frighten*.

Still, it is important to do so, since the type- and token-level agentivity hypothesis are logically independent and make predictions that do not necessarily coincide. The alternation between the reflexive and transitive constructions of the Dutch psych verbs presents an especially useful case study in this regard. This is because it is an alternation that is present both within single verbs and across verbs. For example, contrary to English *to like*, Dutch *verbazen* 'to amaze' may realize its experiencer both in subject and object position, and one may study what determines the choice between both realizations. Since this alternation occurs in several verbs, hypotheses may be put forward as to which verbs will have a stronger preference for an experiencer-subject realization than other verbs. As a consequence, differentiating between the operationalizations of the type- and token-level agentivity hypotheses will be straightforward (see Section 3.1).[11]

The distinction between type and token level should not be confused with the discussion between phrasal and lexical approaches to argument realization (see, e. g., Croft 2003; Kay 2005; Kay and Michaelis 2012; Müller 2006; Boas 2008a, 2008b; Müller and Wechsler 2014). The type and token level are merely two ways in which agentivity – or some other aspect of meaning or function – may have an influence on argument realization. The phrasal and lexical approaches, on the other hand, are two entirely different ways of structuring a theory of argument realization. In the phrasal approach, argument realization is handled by fully abstract argument constructions, while in the lexical approach, argument realization is determined in the lexical entries of verbs. Here, several lexical entries of the same verb may be linked to each other through lexical rules. At the same time, researchers advocating the phrasal approach are perhaps more likely to stress the token level, because it entails that argument constructions exist and carry meaning independently of verbs (Goldberg 1995: 31–42, 2006: 220–221; van Trijp 2015), while proponents of the lexical approach might be more inclined to emphasize the type level because it underlines the role of the verb (Müller 2006; Müller and Wechsler 2014). Within construction grammar, both the phrasal and the lexical approach are possible approaches and are being practiced (for overviews, see Boas 2014: 90–93 and van Trijp 2015: 615–620).

---

11 See Barðdal (2001) and Grafmiller (2013: 166–202) for other such alternations in which both levels can be untangled.

3 Hypotheses

After sketching the research background on psych verbs (Section 2.1) and detailing the distinction between type and token level (Section 2.2), we can turn to the corpus study. In this section, four hypotheses will be presented and operationalized.

3.1 The type-level agentivity hypothesis

While it has been seen that type-level agentivity and token-level agentivity are difficult to distinguish theoretically (Section 2.2), the two levels are, fortunately, easier to discern when being operationalized in corpus research. For that purpose, we make use of two operationalizations employed in earlier corpus studies.

The operationalization of the type-level agentivity hypothesis is taken from Van de Velde (2004: 53–55), who builds on van der Horst (1985). In order to assess to which extent the lexical meaning of a verb presupposes an agentive experiencer, Van de Velde makes use of three of Lakoff's features (1977: 244), viz. *control, volition*, and *responsibility*. He applies these features to the prototypical experiencer of the verbs under scrutiny in his corpus study, i.e., Middle Dutch *(ge/be)denken* 'to think', *(ver)wonderen* 'to wonder', and *ont-/gebreken* 'to miss', as presented in Table 1.[12]

Table 1: Features of agentivity, attributed to the prototypical experiencers of three Middle-Dutch verbs (taken from Van de Velde 2004: 54).

	(ge/be)denken	(ver)wonderen	ont-/gebreken
responsibility	+	+ /-	-
volition	+	+ /-	-
control	+ /-	-	-

The verbs taken up in the present corpus study are *ergeren* 'to annoy', *interesseren* 'to interest', *storen* 'to disturb', and *verbazen* 'to amaze'.[13] In assigning agentivity features to these four verbs, we did not rely on our own judgment, but

---

12 The feature *source* or *causation* could not be employed, as it is anything but clear whether this is a feature of the experiencer or the stimulus (Van de Velde 2004: 54).
13 These verbs were chosen for practical reasons, as they yielded a sufficiently high number of occurrences in the corpora. Originally, the verbs *irriteren* 'to irritate', *verheugen* 'to rejoice', and

consulted five of our colleagues, all native speakers of Dutch. We asked them each to assign the agentivity features employed by Van de Velde (2004) to the prototypical experiencers of the verbs in question. In addition, we asked them to intuitively rank the verbs according to the degree of agentivity of their experiencer.

As expected, there was disagreement about the attribution of agentivity features. Still, from the individual sets of agentivity features, it transpired that that *verbazen* 'to amaze' consistently implied the least agentive experiencer and *interesseren* 'to interest' the most agentive experiencer. This was confirmed by the intuitive rankings. For the scope of this study, only this relative ranking of the verbs from least to most agentive experiencer will be considered, not the precise attribution of agentivity features.

As such, we expect to find *verbazen* 'to amaze' in the transitive construction most often, as it was indicated to imply the least agentive experiencer. Next on the scale of preference from the transitive construction to the reflexive construction should be either *ergeren* 'to annoy' or *storen* 'to disturb'. Finally, *interesseren* 'to interest' is hypothesized to most often appear in the reflexive construction, since it was marked as implying the most agentive experiencer. The type-level agentivity hypotheses will thus be operationalized through the variable *Verb* (see Section 4.2).

3.2 The token-level agentivity hypothesis

The operationalization of the token-level agentivity hypothesis has been informed by Levin and Grafmiller (2012).[14] This operationalization can be used to detect differences in agentivity between individual occurrences, rather than between verbs. Levin and Grafmiller (2012) measure agentivity through animacy, assuming that animate participants are generally more agentive than inanimate participants. This is quite an uncontroversial assumption, as animacy is often pivotal in definitions of agentivity (see, e. g., Lakoff 1977; Langacker 1991: 322). As the experiencer is almost exclusively animate in the corpus data and its animacy would therefore not constitute a useful variable, only the animacy of the stimulus will be examined. It is expected that animate stimuli will elicit the use of the transitive construction, whereas inanimate stimuli will prefer the reflexive construction.

---

verwonderen 'to surprise' were examined as well, but these verbs yielded an insuffient number of occurrences for further analysis.

14 Variations on this operationalization are also used in Grafmiller (2013: 169–174) and Speelman and Geeraerts (2009).

Table 2: Corpus examples of the levels of the variable *Stimulus–Animacy*.

Category	Corpus examples
animate	*wij* 'we' (16), *Tony Herreman* (17), *de grote Israëlische delegatie* 'the large Israeli delegation' (18)
concrete	*het wrak* 'the wreckage' (19), *auto's* 'cars' (20), *de hoge kostprijs van het apparaat* 'the high cost of the machine' (21)
event	*het breken met de regel* 'the breaking of the rule' (21), *ned.-belgië* 'the Netherlands vs. Belgium' (23), *het gegroet* 'the greeting' (24)
abstract	*de landelijke politiek* 'national politics' (25), *de inefficiëntie* 'the inefficiency' (26), *die hypocrisie* 'that hypocrisy' (27)
inanimate residual category	*iets* 'something' (28), *dit soort dingen* 'this kind of stuff' (29), *wat* 'what' (30)
proposition	*dat de afgelaste wedstrijden zo laat worden ingehaald* 'that the canceled games are rescheduled so late' (31), *dat de inhoud ervan bij "buitenstaanders" bekend was* 'That its content was known to "outsiders"' (32), *dat ik daar moeilijkheden mee krijg* 'that that will get me into trouble' (33), *het feit dat we twee goals weggeven uit eigen fout* 'the fact that we give two goals away, which were our own fault' (34)

As in Levin and Grafmiller (2012), animacy is not rated in terms of a binary opposition, but rather in terms of a more detailed scale, ranging from animate persons to abstract entities. This scale is represented in Table 2, along with some corpus examples illustrating each category. The full sentences in which these examples were found are listed in the Appendix.[15]

Our coding of the variable *Stimulus–Animacy* differs from Levin and Grafmiller (2012) only in that we have not distinguished between humans and animate non-humans, and that we have added the *inanimate residual category*. We have not included a separate category of animate non-humans, because this category contained too few occurrences to be truly useful. The label *inanimate residual category* was used when it was evident that the stimulus was inanimate, yet not entirely clear whether it referred to a concrete thing, an event, or an abstract entity. As such, it functioned more or less as a residual category for the inanimate stimuli, hence its name.

---

15 Following Levin and Grafmiller (2012) and Grafmiller (2013), the category *proposition* was also applied to stimuli as in (34) (see Appendix). Here, and in instances such as (33), the category *proposition* differs from the category *subordinate clause* of the variable *Stimulus-Topicality* (see Section 4.2), to which instances such as (33) and (34) do not belong.

3.3 The etymology hypothesis

The third hypothesis is informed by Klein and Kutscher (2005) and concerns the etymology of the verbs.[16] Klein and Kutscher (2005: 20) argue that the argument realization of psych verbs does not depend on their psychological meaning, but that it is determined by their (bond with a former) physical meaning: if a physical meaning is still present in certain uses of the verb, it is this physical meaning that is fed into the linking system. If no physical meaning is present, it is sought in the etymology of the verb (Klein and Kutscher 2005: 41–45). The linking system employed is an instantiation of the type-level agentivity hypothesis, in particular Dowty's (1991) proposal, as modified by Primus (2002). Note that, even for those who do not accept the work by Klein and Kutscher (2005), the etymology hypothesis may still hold intuitive appeal. It only assumes that when the meaning of a verb changes, its argument realization may not (immediately) change with it.

In line with this view, we set out to find an (original) physical meaning for each of the four verbs under scrutiny. To do this, we primarily employed the *Woordenboek der Nederlandse taal* (Dictionary of the Dutch language, de Vries and te Winkel 1998), the *Etymologisch woordenboek van het Nederlands* (Etymological dictionary of Dutch, Philippa et al. 2009), and the *Van Dale groot woordenboek van de Nederlandse taal* (Van Dale Dictionary of Dutch, den Boon and Geeraerts 2005), but we also examined and compared lemmata from Pijnenburg and Schoonheim (2009), Pijnenburg (2001), Verwijs and Verdam (1991), and Lasch et al. (1956).

The psychological meaning of *storen* 'to disturb' seems to have developed from the physical meaning 'to destroy', as in German *zerstören*. In fact, this meaning is still present is some uses, e. g., with respect to bird's nests and certain types of connections, such as radio or wifi connections. In these physical uses, the transitive construction is used categorically, as was the case when the verb was still limited to its original physical meaning 'to destroy'.

The psychological meanings of *ergeren* 'to annoy' and *verbazen* 'to amaze' have developed from the physical meanings 'to damage (lit. *worsen*) something' and 'to make someone act or speak senselessly', respectively.[17] Unlike *storen*,

---

16 Klein and Kutcher do not consider the reflexive construction under scrutiny in the present study (2005: 27), and do not make any direct predictions regarding probabilistic preferences of verbs. As such, the claims that are made in this section are our own, and should not be considered simple reformulations of their work.

17 It is hard – or even principally impossible – to pinpoint exactly when a verb loses its physical meaning and adopts a psychological one, since such semantic changes are gradual

these two verbs have lost their physical meaning in Present-day Dutch. In these meanings, the transitive construction was employed.

The verb *interesseren* 'to interest' is said to be borrowed from the French *intéresser* 'to interest', at a moment when the French verb had already fully acquired its psychological meaning. It never seems to have carried a physical meaning in the history of Dutch. Tracing its development beyond the Dutch language leads us to the Latin *interesse*, which has the physical meaning 'to be in-between'.

In line with the definitions of Primus (2002), Dowty (1991: 572), or Lakoff (1977: 244), or, for that matter, any other linking mechanism based on an instantiation of the type-level agentivity hypothesis, we believe it is safe to say that of these original physical meanings, 'to destroy' presupposes the most agent-like stimulus and the most patient-like experiencer (see also Beavers' affectedness hierarchy, Beavers 2011: 358). In addition, *storen* seems to have, of all four verbs, the most solid bond with its physical meaning, as it is still present in some uses in Present-day Dutch. As such, we expect *storen* to appear most often in the transitive construction.

The original physical meanings of *ergeren* and *verbazen*, 'to damage' and 'to make someone act or speak senselessly', respectively, also imply a rather agent-like stimulus and patient-like experiencer (Lakoff 1977: 244; Dowty 1991: 572). The link with these physical meanings is comparatively weak, however, since they are no longer present in contemporary Dutch. As such, we expect these verbs to appear less often in the transitive construction than *storen*. It is hard to determine whether the stimulus/experiencer of 'to damage' is more agent- or patient-like than the stimulus/experiencer of 'to make someone act or speak senselessly'. This seems to depend on the employed linking mechanism, and the weight attributed to the agentivity features (Lakoff 1977: 244; Dowty 1991: 572, 574; Hopper and Thompson 1980; Croft 1993; Primus 2002, see Section 2.1). Also, the bond both verbs have with their original physical meaning seems equally weak (see footnote 18). Hence, we shy away from making any predictions regarding their preference relative to one another.

Of all the original physical meanings, the meaning 'to be in-between' of *interesseren* seems to imply the least agent-like stimulus and least patient-like

---

by nature (Paul 1898: 74–105). Still, as a working definition, it was assumed that a verb has acquired a psychological meaning when the action or mental state expressed by it can be caused without any physical activity from the part of the stimulus, and may have no physical effects itself – or only indirectly. In Present-day Dutch, this is of course the case for all of our four verbs.

experiencer.[18] The meaning itself ranks low in Transitivity (Hopper and Thompson 1980). In addition, since the verb never carried any physical meaning in the history of Dutch, it can be viewed as having the weakest bond with its original physical meaning of any of the four verbs under scrutiny. Accordingly, there seems less reason for it to prefer the transitive construction than for *storen, ergeren*, and *verbazen*.

To summarize, the etymology hypothesis has led us to suspect that *storen* 'to disturb' will be used most often in the transitive construction, followed by either *ergeren* 'to annoy' or *verbazen* 'to amaze', and finally by *interesseren* 'to interest'. Like the type-level agentivity hypothesis, the etymology hypothesis thus also makes a prediction – be it a different one – regarding the relative preferences of the verbs. As such, the type-level agentivity and etymology hypotheses are both operationalized by the variable *Verb* (see Section 4.2). Because of this, the variable may well be considered our most important variable from a theoretical perspective.

3.4 The topicality hypothesis

The fourth hypothesis is the topicality hypothesis, which concerns the information structure of the sentence. It is based on the finding that long constituents containing new information are generally placed towards the end of a sentence, while topical information is placed at the start (Behaghel 1909; Langacker 1991; Wasow 2002).

In the transitive construction, the stimulus is placed in subject position, while the experiencer occupies the position of direct object. In the reflexive construction, on the other hand, the experiencer is assigned to subject position and the stimulus to the position of prepositional object. In Dutch, as in English, subjects typically precede objects (Haeseryn et al. 1997; Balk-Smit Duyzentkunst 2000; Klooster 2001). As such, it is expected that when the stimulus carries low informational weight and the experiencer high informational weight, the transitive construction is preferred, and conversely, that the reflexive construction is preferred. This hypothesis is operationalized through the variables *Stimulus-Topicality* and *Experiencer-Topicality* (see Section 4.2).

The topicality hypothesis is quite uncontroversial, and is confirmed in nearly all corpus studies concerning argument realization (e. g., Colleman 2006; Bresnan et al. 2007; Levin and Grafmiller 2012). As such, we can be fairly confident that it will not be refuted in the present study. Still, this hypothesis may present itself as an alternative to the token-level agentivity hypothesis: it would be interesting to

---

18 From the dictionary entries, it was not clear whether the subject of Latin *interesse* should be considered to correspond to the stimulus or experiencer of Dutch *interesseren*. However, this question seems anachronistic.

investigate whether the choice between a reflexive and transitive construction is more strongly determined by meaning or by information structure.

4 Data

4.1 Corpora and composition of the dataset

In order to obtain a representative sample of Dutch language use and to accumulate a reasonable amount of data, we have made use of the *ConDiv* corpus and the *Corpus of Spoken Dutch* (Corpus Gesproken Nederlands or CGN).[19] Both these corpora contain language material from around the year 2000, and provide a broad range of language genres, from written and spoken language respectively. The ConDiv corpus comprises chat, email, and newspaper language from Belgium and the Netherlands, totaling about 45 million words. The CGN contains transcripts ranging from spontaneous conversations to pre-pared lectures, also from Belgium and the Netherlands, and consists of nearly 10 million words. More information about the ConDiv and CGN corpora can be found in Grondelaers et al. (2000), Oostdijk et al. (2002), and van Eerten (2007).

From these corpora, all hits of the verbs *ergeren* 'to annoy', *interesseren* 'to interest', *storen* 'to disturb', and *verbazen* 'to amaze' were extracted using AntConc software (Anthony 2011). Not all attestations were retained, however. First, 25 occurrences of *interesseren* were excluded whose meaning differs from both the transitive and reflexive constructions, as in (5).

(5)   *Op termijn hoopt GroenLinks een aantal van deze vrouwen te*
       on term hopes GroenLinks a number of these women to
       interesseren voor raadswerk.
       interest for council_work
       'In the long run, GroenLinks hopes to motivate a number of these women for council work.'

Second, the dataset contained two hits in which a free relative clause functioned as the experiencer, as in (6). Since two hits was considered too limited to warrant a separate level for the variable *Experiencer-Topicality* (see Section 4.2), they were removed from the dataset.

---

19 *Corpus of Spoken Dutch*: http://tst-centrale.org/nl/tst-materialen/corpora/corpus-gesproken-nederlands-detail; *ConDiv corpus of written Dutch*: http://neon.niederlandistik.fu-berlin.de/static/digitaal/digitaal-11.html. Since the small diachronic component and the Bulletins of Acts, Orders and Decrees from the Netherlandic and Belgian governments in the *ConDiv* corpus yielded too few hits to be useful, these sections of the corpus were not used.

(6)   *Wie de schrijver kent,   kortom, zal het niet verbazen dat de*
        who the writer   knows in-short will it   not amaze   that the
        P.C. Hooftprijs 1998   wordt uitgereikt tijdens een besloten bijeenkomst.
        P.C. Hooft-award 1998 is   awarded during a   closed   meeting
        'Those who know the writer, in short, will not be surprised to hear that the
        P.C. Hooft award 1998 is awarded during a closed meeting.'

Third, the occurrences of *storen* 'to disturb' in which a clear physical meaning was present were not included, since these do not allow for a reflexive construction and are not examples of psychological verbs to begin with. Included here are physical meanings in which the experiencer was an inanimate object, as in (7) (also see Section 3.3), or instances in which *storen* was used in the meaning of 'to interrupt someone, who is working or talking to someone else, by asking for their attention', as in (8). These cases, a total of 201, could be distinguished quite easily from instances with psychological meaning, as in (9).

(7)   *met pods kan de F16-piloot de aanvallers en   de vijandige*
        with pods can the F16-pilot   the attackers   and the enemy
        radar actief storen.
        radar active disturb
        'With pods, the F16-pilot can actively deflect enemy missiles and jam the enemy radar.'

(8)   *Goedenmiddag, met softwarehouse   Been, sorry dat ik u*
        good-afternoon with software-house Been   sorry that I   you
        stoor,   maar ik wil   het even   met u
        disturb but   I   want it   briefly with you
        over   het millennium probleem hebben.
        about the millennium problem   have
        'Good afternoon, this is the software company Been, sorry to bother you, but I would like to talk to you briefly about the millennium problem.'

(9)   *Soms   stoort   het hem, al die aandacht voor junks.*
        Sometimes disturbs it   him   all that attention for   junks
        'Sometimes it disturbs him, all that attention for junks.'

Fourth, all participles were excluded: many are lexicalized as adjectives with a specific meaning, such as *gestoord* 'insane' (10). It was therefore concluded that individual inspection of these hits would not yield sufficient useful hits for the inspection to be carried out.

(10) <lonely_r> *Ja* *ik* *ben nu* *officieel gestoord verklaard*
      <lonely_r> yes I   am now officially insane   declared
      <masterm> *nu   pas?*
      <masterm> now just
      '<lonely_r> Yeah, I'm now officially declared insane. <masterm> Only now?'

Lastly, as in Levin and Grafmiller (2012: 25), all 291 occurrences in which the experiencer or stimulus were not expressed were excluded as well, because in Dutch the subject is obligatory (Haeseryn et al. 1997: 1131; van der Horst 2008: 163–164, 566). As such, if there is no experiencer or no stimulus present, the construction that would realize the absent participant in subject position is not possible. For example, using the transitive variant in (11) would require realizing the stimulus in subject position, yet this stimulus is simply not present. Likewise, the absence of an overtly expressed experiencer in (12) makes it impossible to express the same proposition using the reflexive variant, at least without inserting a dummy experiencer.

(11) *nou eerlijk waar nou dan verbaas je je elke keer*
      now honest truly now then amaze you you every time
      'Now, honestly, then you're amazed time and again.'

(12) *over   welke informatie   er   moet staan*, *en op welke plaats*
      About which information there must stand and on which place
      om   zo min mogelijk te storen
      in_order_to so least possible to disturb
      'about which information should be there, and where it should be placed in order to disturb as little as possible.'

The remaining dataset still contained 2785 occurrences. Among these were some occurrences of the verb *ergeren* 'to annoy' which could not be straightforwardly classified as reflexive or transitive, because the reflexive pronoun is absent; an example is (13). Because the experiencer occupies subject position and the stimulus is realized as a prepositional object, these occurrences can still be regarded as a reflexive construction missing its reflexive pronoun.[20] Since they only appeared five times in the dataset, these hits were grouped together with the reflexive result.

---

20 In three of these occurrences, one of which is (11), the experiencer was the unstressed second person singular *je* 'you', which is formally identical to the 'missing' reflexive pronoun. As such, these occurrences could present instances of *horror aequi* (Rohdenburg 2003: 205).

(13) *Niet naar kijken*, *dan kan je   er   ook niet aan ergeren*.
      Not at   look   then can you there also not to   annoy
      'Don't look at it, then it cannot annoy you.'

4.2 Variables

To this dataset, the following variables with their respective levels were added manually (where necessary) and automatically (where possible). The response variable is the constructional variant: all occurrences in the dataset instantiate either the transitive or the reflexive construction. This means that when we state that the probability of using the transitive construction goes down, the probability of using the reflexive goes up, and vice versa. Among the explanatory variables, we discern between the hypothesis-driven variables, which represent the operationalization of the hypotheses discussed in Section 3, and the nuisance variables, for which no hypotheses have been proposed.[21] Still, it is important to also check for the possible influence of such nuisance variables, since they might inflate the influence of a hypothesis-driven variable (a false positive) or mask an authentic influence (a false negative). For instance, it might be the case that animate stimuli are more prevalent in informal language use, and that the transitive variant is the informal one, without animate stimuli directly eliciting the use of the transitive construction. If, in such a case, the variable *Register* is not taken into account, we might be led to conclude – erroneously – that there is a significant influence of animacy on transitivity. Still, from a theoretical standpoint, these nuisance variables are of secondary importance, and are therefore presented only succinctly.

Response variable	*Levels*
– Variant:	*transitive, reflexive*

Explanatory variables
– Hypothesis-driven variables

– Verb:	*ergeren, interesseren, storen, verbazen*
– Stimulus-Animacy:	*animate, inanimate, concrete, event, abstract, proposition*
– Stimulus-Topicality:	*first person, second person, third person-pronoun, definite noun, indefinite noun, subordinate clause*

---

21 These nuisance variables were chosen because (i) they represent unproblematic classifications of the data (e. g., there is little discussion about whether a particular stimulus is singular or plural), and (ii) they can be easily added to the dataset.

    –   Experiencer-Topicality: *first person, second person, third person-pronoun, definite noun, indefinite noun*

–   Nuisance variables
    –   Stimulus-Number: *singular, plural*
    –   Experiencer-Number: *singular, plural*
    –   Negation: *with, without*
    –   Finiteness: *finite, non-finite*
    –   Tense/Mood: *present, past, future, conditional*
    –   Country: *Belgium, the Netherlands*
    –   Register: *chat, informal speech, formal speech, email, mass newspaper, quality newspaper*

The variable *Verb* operationalizes the type-level agentivity hypothesis and the etymology hypothesis, which make conflicting predictions regarding the preferences of the verbs for the reflexive or transitive construction. The variable *Stimulus-Animacy* operationalizes the token-level agentivity hypothesis.

    Stimulus-Topicality and *Experiencer-Topicality* offer two operationalizations of the topicality hypothesis. The first- and second-person pronouns bear the lowest informational weight, since they are directly present in the context. Third-person pronouns bear slightly heavier informational weight, as they are not necessarily present in the context, but are still familiar enough to elicit pronominal use. For the nouns, a distinction is made between definite and indefinite nouns, with indefinite nouns contributing newer information. Finally, subordinate clauses are considered heaviest in terms of informational weight.

    In the description of Dutch word order, a central position is attributed to the verbal poles (also called *Klammern* in the German tradition, see Zifonun et al. 1997: 1498; Zwart 2011: 26–37). Canonical Dutch main clauses put the finite verb in second position (so-called V2-sentences), while yes-no questions and conditional clauses without conjunctions place the verb in first position (so-called V1-sentences). In these sentences, the first pole is said to contain the finite verb, and the second pole all other non-finite verbs. In subordinate clauses, the finite verb is placed at the end of the sentence (so-called Vf-sentences). Here, the subordinator is said to occupy the first pole, while all verbs occupy the second pole. These two poles divide the sentence into three fields, viz. prefield, middle field, and postfield, as exemplified in Table 3.[22]

---

22 This description in terms of "poles" and "fields" is called the topological field approach (Zwart 2011: 26–37).

Table 3: Topological fields in Dutch (partially adopted from De Smet and Van de Velde 2013: 537).

V2 (i)	Prefield	1st pole	Middle field	2nd pole	Postfield
(a)	*Ik*	*heb*	*dat book in de VS*	*gelezen*	∅
	I	have	that book in the US	read	
(b)	*Ik*	*heb*	*dat boek*	*gelezen*	*in de VS*
	I	have	That book	read	in the US
(c)	*Dat boek*	*heb*	*ik in de VS*	*gelezen*	∅
	That book	have	I in the US	read	
(d)	*In de VS*	*heb*	*ik*	*gelezen*	*dat het boek uitverkocht is*
	In the US	have	I	read	that the book sold_out is

Vf (ii)	Prefield	1st pole	Middle field	2nd pole	Postfield
(a)	∅	*dat*	*ik dat boek*	*heb gelezen*	*in de VS*
		that	I that book	have read	in the US
(b)	∅	*dat*	*ik dat boek in de VS*	*heb gelezen*	∅
		that	I that book in the US	have read	

Unless another constituent is specifically topicalized, a subject in Dutch will occupy the prefield or, in V1- and Vf-sentences, the first position of the middle field, immediately after the first pole (see Table 3). The direct object is placed in the middle field, unless it is specifically topicalized (as in (ic) in Table 3). Neither subject nor direct object can appear in postfield position, unless they are realized as subordinate clauses, as in (id). The prepositional object has a wider range of positions available to it. It can appear both in middle field and post field, and, if topicalized, in prefield. Still, Willems and De Sutter (2015, 2016) have convincingly shown that its standard position is the postfield.

The reflexive construction assigns the experiencer to subject function and the stimulus to the function of prepositional object. Hence, we expect this construction to be preferred if the experiencer is informationally light and the stimulus informationally heavy, as in (14). The transitive construction, on the other hand, realizes the stimulus in subject function and the experiencer in direct object function. Therefore, we expect to see this construction being used if the stimulus is light and the experiencer is heavy, as in (15). We make one exception to this prediction, namely when the stimulus takes the form of a subordinate clause.[23] In that case, both subject and prepositional object can be easily placed in postfield position. As such, the topicality hypothesis as

---

23 This is only relevant for the stimulus, since the dataset no longer contained any experiencers realized as subordinate clauses (see Section 4.2).

stated here makes no predictions concerning the behavior of stimuli realized as subordinate clauses.

(14) *IK stoor me aan al te chauvinistisch gedrag.*
 I disturb me to all too chauvinistic behavior.
 'Behavior that is too chauvinistic disturbs ME.'

(15) *Oh ja, dat interesseert veel mensen.*
 Oh yes that interests many people
 'Oh yes, a lot of people find that interesting.'

The nuisance variables *Stimulus-* and *Experiencer-Number, Negation,* and *Country* are self-explanatory. *Finiteness* distinguishes between the finite and non-finite forms of the verbs. For the variable *Tense/Mood*, the present perfect was coded as *past*. The ordinal variable *Register* contained several levels in a clear ordering, from (highly) informal to formal. The *ConDiv* corpus supplied material for the levels chat conversations, emails, mass newspapers, and quality newspapers. Data for the levels informal and formal speech were provided by the CGN-corpus (based on the division presented in Plevoets 2008: 80).

5 Analysis

The dataset, as presented in the previous section, was analyzed by means of logistic regression. Regression analysis is a statistical technique which measures the effect of one or several explanatory (or *independent*) variables on a single response (or *dependent*) variable. In essence, it is simply a mathematical function which allows the researcher to enter values for a number of parameters that contain information on the explanatory variables. The function then returns a predicted value for the response variable. If the response variable is binary, as is the case in the present study, the logistic function is used. This function can be visualized as describing a hyperplane in a Cartesian space, with the different parameters and the response variable as its dimensions. When fitting this function to the data, the estimates of the function are chosen so as to minimize the distance between this hyperplane and the observed values.

Numeric explanatory variables can be implemented directly as parameters in a regression model. Categorical variables require an additional step for their information to be encoded in the regression model. Here, we employ dummy coding, an approach in which one of the levels of the categorical explanatory variable – for technical reasons, this is often the most frequent level – is

arbitrarily chosen as the reference level. The remaining levels are implemented as separate parameters which can take the value 0 or 1. For instance, for the variable *Verb*, the level *interesseren* is chosen as a reference level. The variable *Verb* is then implemented as three parameters in the regression model. If, in one data point, the verb *ergeren* is used, the first parameter of these three is set to 1, and the other two parameters to 0. If the verb *storen* is used, the second parameter is set to 1, and the other two to 0. If the verb *verbazen* is used, the third is set to 1, and the others to 0. If the verb *interesseren*, i. e., the reference level, is used, all parameters are set to 0. A categorical variable with *n* levels is thus implemented through *n-1* parameters. The same is done for the binary response variable, which is implemented as a single parameter. This means that one level is selected as the reference level and is assigned value 0, while the other is named the success level, and receives value 1.

When working with several explanatory variables, regression analysis allows the researcher to assess the influence of a single explanatory variable while keeping the other explanatory variables constant. It is thus statistically superior to simple bivariate testing, in which the influence of one explanatory variable is measured while ignoring the effects of another, a procedure that can lead to problematic results (Speelman 2014: 499–502). As such, it provides a very useful tool to investigate linguistic alternations in which the choice of the language user is simultaneously determined by several factors, and it has become the work-horse technique in the field of alternation studies (e. g., Bresnan et al. 2007; Speelman and Geeraerts 2009; Grafmiller 2013; Pijpops and Van de Velde 2014).[24]

The crucial task in a logistic regression analysis is to determine which explanatory variables to include in the regression model. To do this, we have made use of the bidirectional stepwise variable selection procedure described in Venables and Ripley (2002: 173–177), which was run automatically in R (R Core Team 2014).[25] The model which was returned by this procedure was tested for the following criteria: firstly, each variable had to significantly improve the model; otherwise, it would be dropped. Secondly, as a rule of thumb, no more parameters were included in the model than the number of occurrences of the least frequent variant divided by 20 (Speelman 2014: 530). The residual deviance

---

24 For introductions to logistic regression analysis, see Baayen (2008), Gries (2013), and Speelman (2014). For more advanced mixed models, see Gries (2015).

25 We have used R version 3.1.1. The following R-packages were employed in the analysis and presentation of the data: caret (Kuhn 2008), effects (Fox 2003), extrafont (Chang 2014), lme4 (Bates et al. 2013), MASS (Venables and Ripley 2002), party (Hothorn et al. 2006; Strobl et al. 2007, 2008) and rms (Harrell 2013).

was not allowed to be much higher than the degrees of freedom, and the Hosmer-Lemeshow-Cessie goodness of fit test could not return a significant p-value, which would indicate a significant lack of fit (Speelman 2014: 528). Lastly, multicollinearity needed to be below threshold levels. Multicollinearity was not found to be a problem, as all variance inflations factors were well below 4 (see the Appendix, Speelman 2014: 528).

Table 4 presents the resulting logistic regression model. The AIC-value is an indication of the extent to which the model was able to explain the variation. The C-index indicates the model's predictive quality. A C-index of 0.800 or higher is taken to indicate a model of reasonable quality (Speelman 2014: 515), a standard which, demonstrably, is easily met.

The success level of the response variable was the reflexive construction. The estimates and their confidence intervals are rounded off to two decimals, the p-values to four. The scale of the estimates is the so-called logit scale.[26] Positive estimates for regressors indicate an increase of the probability of the reflexive construction, relative to the reference level; negative estimates indicate a decrease. For instance, the positive estimate for *ergeren* 'to annoy' indicates that the probability of occurrence of the reflexive construction with *ergeren* is increased relative to its occurrence with *interesseren* 'to interest'.

Table 4: Regression model predicting the use of the reflexive and transitive construction.

− AIC: 2072	− Total number of hits: 2785			
	− Hits of the transitive: 1911			
− C-index: 0.904	− Hits of the reflexive: 874			

Explanatory variable	Level	Estimate	Confidence interval		P-value
			2.5 %	97.5 %	
	intercept	−3.78	−4.27	−3.30	<0.0001
Verb	*interesseren*				Reference level
	ergeren	4.00	3.60	4.41	<0.0001
	storen	2.14	1.80	2.48	<0.0001
	verbazen	1.26	0.90	1.64	<0.0001

(*continued*)

---

26 As for what this logit technically signifies and how to calculate the probability of each variant from this logit, see Speelman (2014: 493–498).

Table 4: (*continued*)

Stimulus-Topicality	*third person pronoun*				Reference level
	first person	0.42	−0.49	1.32	0.3600
	second person	−0.05	−1.44	1.10	0.9342
	definite noun	1.80	1.45	2.16	<0.0001
	Indefinite noun	1.97	1.51	2.45	<0.0001
	subordinate clause	0.59	0.18	1.00	0.0048
Experiencer-Topicality	*first person*				Reference level
	second person	0.69	0.31	1.07	0.0003
	third person pronoun	1.40	1.08	1.71	<0.0001
	definite noun	1.38	1.02	1.75	<0.0001
	indefinite noun	1.34	0.83	1.86	<0.0001
Country	*Belgium*				Reference level
	the Netherlands	0.88	0.64	1.13	<0.0001
Stimulus-Animacy	*proposition*				Reference level
	animate	−1.37	−1.91	−0.84	<0.0001
	inanimate	0.34	−0.11	0.79	0.1404
	concrete	−0.08	−0.54	0.37	0.7163
	event	−0.63	−1.17	−0.10	0.0198
	abstract	0.94	0.44	1.44	0.0002
Tense	*present*				Reference level
	past	0.20	−0.11	0.51	0.1967
	future	−0.82	−1.48	−0.21	0.0112
	conditional	−1.60	−2.42	−0.89	<0.0001
Negation	*without*				Reference level
	with	−0.61	−0.89	−0.35	<0.0001
Stimulus-Number	*singular*				Reference level
	plural	0.68	0.32	1.04	0.0002
Register	*e-mail*				Reference level
	chat	−0.22	−0.64	0.18	0.2810
	informal speech	0.02	−0.42	0.45	0.9335
	formal speech	−0.29	−0.67	0.09	0.1331
	mass newspaper	−0.69	−1.03	−0.36	<0.0001
	quality newspaper	−0.35	−0.78	0.07	0.1057

The effect plots in Figure 1 provide a visualization of the influence of the variables on the choice of variant. Only the effect plots of the hypothesis-driven variables are shown, since they are of primary concern to our study and are the most important variables in the regression model (see Figure 2). The y-axis shows the estimated probability of the reflexive construction for each level, with the values of the other explanatory variables adjusted to the mean (Buis 2007). This means they represent the effect of each of the explanatory variables, while simultaneously taking the other variables into account. The "whiskers" indicate confidence intervals. Finally, we ran a random forest analysis on the
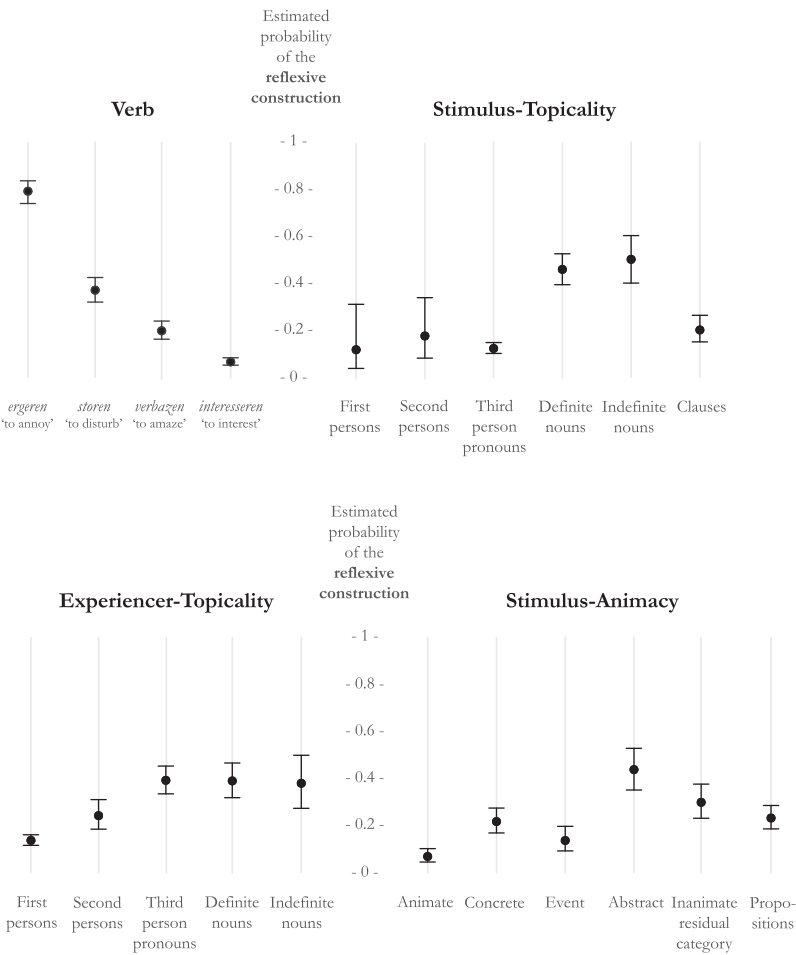

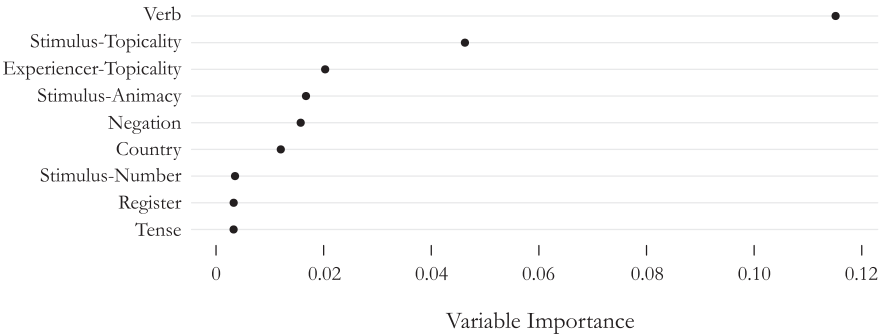**Figure 1:** Effect plots of the hypothesis-driven variables.

Figure 2: The importance of the explanatory variables in the regression model.

model to assess the importance of each explanatory variable in the regression model (for details, see Strobl et al. 2007). The results of this are visualized in Figure 2.

6 Interpretation

Figure 2 indicates that the four hypothesis-driven variables are the most important explanatory variables in the model, although *Experiencer-Topicality* and *Stimulus-Animacy* perform only slightly better than the nuisance variables *Negation* and *Country*. In what follows, these variables are discussed in order of importance in the model. The effect of the nuisance variables seems negligible, except perhaps for *Negation* and *Country*. It appears that the transitive construction is more popular in negated sentences and in Belgium.

6.1 The variable *Verb*: The type-level agentivity hypothesis and the etymology hypothesis

By far the most important explanatory variable in the model is the variable *Verb*. *Ergeren* 'to annoy' most strongly prefers the reflexive argument construction, followed by *storen* 'to disturb' and *verbazen* 'to amaze', and finally *interesseren* 'to interest', which has the strongest preference for the transitive variant. This ranking does not correspond to any of our hypotheses, nor does it change when

we consider the raw numbers rather than at the estimated probabilities in the regression model.[27]

The agentivity hypothesis is thus not confirmed at the type level. As shown below, the observed order of verbs, from most frequent in the transitive construction to most frequent in the reflexive construction, is in fact quite different from the hypothesized order based on the lexical meaning of the four verbs (Figure 3).
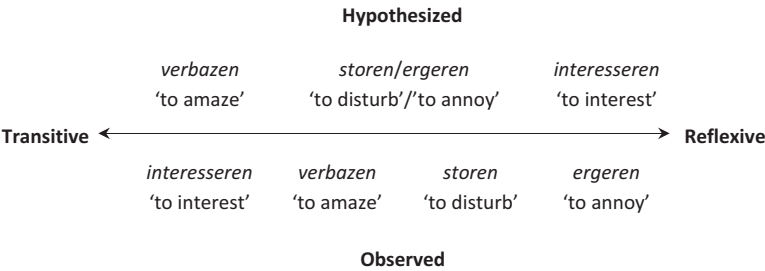
Hypothesized

	verbazen	*storen/ergeren*	*interesseren*
	'to amaze'	'to disturb'/'to annoy'	'to interest'

Transitive ⟵————————————————————⟶ **Reflexive**

	interesseren	*verbazen*	*storen*	*ergeren*
	'to interest'	'to amaze'	'to disturb'	'to annoy'

Observed

Figure 3: Results of the type-level agentivity hypothesis.

It appears, then, that we have not been able to infer the dominant argument construction of a verb from its lexical meaning. That does not mean that it is impossible to do so. For one, we have limited ourselves to predicting probabilistic preferences of verbs for which multiple argument realizations were possible, whereas most linking theories focus on verbs that (almost) categorically appear with only a single argument realization (e. g., Dowty 1991; Baker 1997; Van Valin 2004; see Section 2.1). Moreover, this study only investigated the behavior of four verbs. These four may very well happen to represent the exception rather than the rule. To truly confirm or reject the type-level agentivity hypothesis for the psych verbs, all psych verbs should be taken into account. Our results do show that, even if the agentivity hypothesis were shown to hold true on the type level, it should be considered a general tendency to which exceptions are possible, and not an absolute law.

The variable *Verb* was also used in the operationalization of the etymology hypothesis. This hypothesis was an attempt to predict the relative preferences

---

27 It also does not change when we include the occurrences of the physical meaning of *storen* 'to disturb', although this, of course, does raise the number of transitive occurrences of that verb (see Section 4.1).
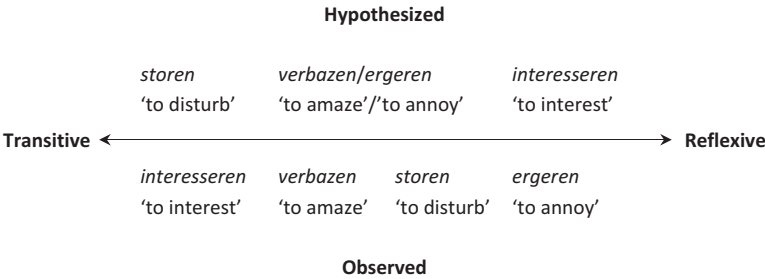
Hypothesized

storen	verbazen/ergeren	interesseren
'to disturb'	'to amaze'/'to annoy'	'to interest'

Transitive ←——————————————————————→ **Reflexive**

interesseren	verbazen	storen	ergeren
'to interest'	'to amaze'	'to disturb'	'to annoy'

Observed

Figure 4: Results of the etymology hypothesis.

of verbs for a particular argument structure based on their etymology, or, more specifically, on their original physical meaning and the link they may still have with that meaning today. Again, the hypothesized order does not correspond to the observed order of verbs. We were thus also unable to predict the preferred argument construction of a verb based on its diachronic development (Figure 4).

6.2 The variables *Stimulus*- and *Experiencer-Topicality*: The topicality hypothesis

As opposed to the variable *Verb*, the variable *Stimulus-Topicality*, the second most important explanatory variable, does show the expected effect, with nouns exhibiting an outspoken preference for the reflexive construction compared to pronouns (see Figure 1). This variable thus clearly confirms the topicality hypothesis. Disregarding the subordinate clauses, for which no predictions were made, we find that informationally heavy stimuli are more likely to be assigned the function of prepositional object in the reflexive construction than the function of subject in the transitive construction.

More surprising is the influence of *Experiencer-Topicality*. The topicality hypothesis predicted that the experiencers would behave exactly opposite from the stimuli, with informationally heavier experiencers preferring the transitive construction to the reflexive construction. However, we find that, just like the stimuli, the informationally heavier experiencers show an increased preference for the reflexive construction, not the transitive one (see Figure 1).

In retrospect, however, this behavior might not be as counterintuitive as it appears at first sight. As far as the stimulus is concerned, the choice between

the transitive and reflexive construction constitutes a choice between subject and prepositional object function. As argued in Section 4.2, prepositional objects can easily be placed in postfield position and are therefore typically associated with a heavy informational weight. Subjects, on the other hand, cannot be placed in postfield position.[28] This means that, in terms of the topicality of the stimulus, the choice between a transitive and a reflexive construction matters greatly. Conversely, with regard to the experiencer, the choice between the transitive and reflexive constructions boils down to a choice between subject or direct object function. Like the subject, the direct object in Dutch cannot be placed in postfield position.[29] For the experiencer, the variants thus constitute a choice between two functions which are both associated with a rather light informational weight. For the experiencer, the choice between transitive and reflexive construction is therefore of less importance than for the stimulus.

Now let us consider the case in which both stimulus and experiencer carry a heavy informational weight, for example, if they are both indefinite nouns. Firstly, the choice for the reflexive construction benefits the stimulus to a greater extent than the transitive construction benefits the experiencer, as argued in the previous paragraph. Secondly, the reflexive pronoun provided by the reflexive construction might function as an informationally light "pause" between a heavy experiencer and stimulus, whereas in the transitive construction, both participants need to be placed close to one another (see Table 3). Additionally, the preposition belonging to the prepositional object in the reflexive construction can also help to enlarge the distance between the heavy experiencer and stimulus. From this, the choice for a reflexive construction follows quite naturally.

For the stimuli, we find the most outspoken difference in preference for the reflexive construction between the pronouns and the nouns. The experiencers, on the other hand, display the largest increase from the first to the third person. This difference in behavior of the stimuli and the experiencers can be accounted for quite easily: nearly 95 % of the stimuli are third persons, rendering it more useful for the language user to employ the transitive-reflexive alternation in order to accommodate for the difference in informational weight between pronouns and nouns. Conversely, more than 80 % of the

---

28 Unless it takes the form of a subordinate clause, for which no predictions were made (see Section 4.2).

29 Again, unless it takes the form of a subordinate clause. These hits, however, were excluded from the dataset (see Section 4.2).

experiencers are pronouns, so a difference in preference is more likely to be found between the persons.

6.3 The variable *Stimulus-Animacy*: The token-level agentivity hypothesis

For the variable *Stimulus-Animacy*, Figure 1 shows that the animate stimuli prefer the transitive construction, whereas the abstract stimuli prefer the reflexive construction. The remaining inanimate levels *proposition*, *inanimate*, *concrete*, and *event* rank in-between, with overlapping confidence intervals. This confirms the prediction based on the token-level agentivity hypothesis. However, of all the hypothesis-driven variables, *Stimulus-Animacy* is the least important variable in the model, barely outperforming the nuisance variables *Negation* and *Country*. The – unexplained – lexical preferences of the verbs and the information structure therefore seem to determine the choice of argument construction more strongly than the animacy of the stimuli.

Still, *Stimulus-Animacy* is only a partial operationalization of agentivity, and while we do believe it is a reliable one, we were not able to measure any possible differences in agentivity between animate stimuli, or between stimuli in the form of an abstract entity.[30] Also, this operationalization only allowed us to take into account the agentivity of the stimulus, not of the experiencer. Therefore caution is in order so as not to directly interpret the relatively low variable importance of *Stimulus-Animacy* in Figure 2 as agentivity playing only a minor role in argument realization. After all, *Stimulus-Animacy* confirms the token-level agentivity hypothesis, even though its influence of may be a little underwhelming.

To round off this section, Table 5 provides a quick overview of the explanatory variables which determine the alternation between the transitive and reflexive argument construction. Only the four most important variables, i. e., the hypothesis-driven variables, are shown. They are ordered from most to least important.

---

30 For instance, if the writer would like to present an animate being as very active in one sentence, and as rather apathetic in another, this difference would not be picked up, as the stimuli in both occurrences would be labelled *animate*.

Table 5: Variables determining the choice between the transitive and reflexive argument constructions, from most to least important.

Variables	Transitive				Reflexive
– *Verb*	←——→				
	interesseren 'to interest'	*verbazen* 'to amaze'	*storen* 'to disturb'		*ergeren* 'to annoy'
– *Stimulus-Topicality*	←——→				
	first persons	second persons	third- person pronouns	definite nouns	indefinite nouns
– *Experiencer-Topicality*	←——→				
	first persons	second persons	third- person pronouns	definite nouns	indefinite nouns
– *Stimulus-Animacy*	←——→				
	animate		other inanimate categories		abstract entity

7 Conclusions

The present study may serve as an example of the importance of distinguishing between the type- and token-level agentivity hypotheses, which are often conflated in the available theoretical and quantitative literature (Dowty 1991: 579–581; Goldberg 1995; Levin and Grafmiller 2012: 21–24; see Section 2.2). While the agentivity hypothesis has been confirmed at the token level, it has not been confirmed at the type level.

This failure to find confirmation for the type-level agentivity hypothesis means that our findings cannot be used in support of theories which attempt to infer the argument construction of a verb directly from its lexical meaning, or in which the argument construction iconically reflects the lexical meaning of a verb, such as Baker's Uniformity of Theta Assignment Hypothesis (Baker 1988, 1997), Dowty's proto-roles (Dowty 1991), Langacker's *flow of energy* (Langacker 1991), Van Valin's Macroroles (Van Valin 2004), etc. Stating that our findings cannot be used in support of such linking theories, however, does not mean that they provide evidence against them. This is firstly because the present study did not directly deal with the problem of psych verbs itself, as explained in Section 2.1. In addition, as stressed in Section 6.1, we only investigated four verbs in a single language, and dealt with probabilistic constraints rather than categorical ones. Finally, absence of evidence is different from evidence of absence.

To truly verify the type-level agentivity hypothesis for the psych verbs, more systematic research is needed, ideally incorporating all (psychological) verbs. Each verb could, for instance, be assigned a score indicating how strongly its lexical meaning presupposes an agent-like experiencer and a patient-like stimulus, or the other way round. These scores can then be correlated with their preference for an argument construction. In what way such scores can be assigned, though, is yet unclear to us. Meanwhile, the confirmation found for the token-level agentivity hypothesis means that our results can, in fact, be used in support of theories claiming that argument realization adds meaning to the utterance (see, e. g., Goldberg 1995: 31–43; but also Dowty 1991: 579–581).

The following reformulates these conclusions in constructionist terms. The confirmation of the token-level agentivity hypothesis indicates that argument constructions do indeed contribute to the compositional meaning of the sentence, independent of the verb or of other lexical elements (Goldberg 1995: 1–21, 31–39). Argument constructions can therefore be said to carry meaning. However, in isolation, our data do not indicate that verbs are more likely to combine with argument constructions whose meaning is more compatible with the lexical meaning of the verb (Goldberg 1995: 43–66).

Like the type-level agentivity hypothesis, the etymology hypothesis as formulated in Section 3.3 also failed to correctly predict the individual lexical preferences of the verbs for the transitive or the reflexive construction. These preferences thus remain unexplained. This is particularly unfortunate, since they contained by far the most important information to predict the argument construction. We hope future research may shed light on this matter.

Lastly, the topicality hypothesis was confirmed, be it with some major revisions. The variables that operationalized this hypothesis successfully predicted the transitive–reflexive alternation. This confirmation, along with the confirmation of the token-level agentivity hypothesis in Dutch, dovetails with the findings of Levin and Grafmiller (2012) and Grafmiller (2013) in English. We can thus attempt to formulate a preliminary interlingual generalization, which we suspect holds for all psychological verbs exhibiting syntactic variation: if the stimulus is more agentive and/or has a lower informational weight, it is more likely to appear in subject position. However, when it comes to the transitive-reflexive alternation in Dutch, it seems that the stimulus and the experiencer are not so much in competition with regard to informational wieght. Rather, when both stimulus and experiencer carry a heavy informational weight, the reflexive construction is preferred to the transitive one.

of our data and helpful advice on the operationalizations of the hypotheses. We also owe thanks to three anonymous reviewers for valuable comments and suggestions. Thanks are also due to Isolde Vande Walle and editor-in-chief Hubert Cuyckens for additional remarks which greatly helped to improve this paper. All remaining errors are, of course, are own.

References

Ackerman, Farrell & John Moore. 2001. *Proto-properties and grammatical encoding: A correspondence theory of argument selection*. Stanford: CSLI Publications.

Alexiadou, Artemis & Gianina Iordăchioaia. 2014. The psych causative alternation. *Lingua* 148. 53–79.

Anthony, Laurence. 2011. *AntConc (Computer Software, version 3.3.3)*. Tokyo: Waseda University.

Arad, Maya. 1996. A minimalist view of the syntax-lexical semantics interface. *UCL Working Papers in Linguistics* 8. 1–30.

Baayen, Rolf Harald. 2008. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Baker, Mark. 1988. *Incorporation: A theory of grammatical function changing*. Chicago: University of Chicago press.

Baker, Mark. 1997. Thematic roles and syntactic structure. In Liliane Haegeman (ed.), *Elements of grammar: Handbook in generative syntax*, 73–137. Dordrecht: Kluwer.

Balk-Smit Duyzentkunst, Frida. 2000. *Grammatica van het Nederlands [Grammar of Dutch]*. Den Haag: Sdu Uitgevers.

Barðdal, Johanna. 2001. The Perplexity of Dat–Nom Verbs in Icelandic. *Nordic Journal of Linguistics* 24(1). 47–70.

Barðdal, Johanna. 2009. The development of case in Germanic. In Johanna Barðdal & Shobhana Chelliah (eds.), *The role of semantic, pragmatic, and discourse factors in the development of case*, 123–159. Amsterdam: John Benjamins.

Bates, Douglas, Martin Maechler, Ben Bolker & Steven Walker. 2013. *lme4: Linear mixed-effects models using Eigen and S4*. R package version *1.0-4*.

Beavers, John. 2011. On affectedness. *Natural Language and Linguistic Theory* 29 (2). Dordrecht. 335–370.

Behaghel, Otto. 1909. Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern. *Indogermanische Forschungen* 25. 110–142.

Belletti, Andriana & Luigi Rizzi. 1988. Psych-verbs and θ-Theory. *Natural Language and Linguistic Theory* 6(3). 291–352.

Bennis, Hendrik Johannes. 1986. *Gaps and dummies*. Tilburg: Tilburg University PhD dissertation.

Bennis, Hendrik Johannes. 2004. Unergative adjectives and psych verbs. In Artemis Alexiadou, Elena Anagnostopoulou & Martin Everaert (eds.), *The unaccusativity puzzle: Explorations of the syntax-lexicon interface*, 84–113. Oxford: Oxford University Press.

Boas, Hans. 2008a. Determining the structure of lexical entries and grammatical constructions in Construction Grammar. *Annual Review of Cognitive Linguistics* 6(1). 113–144.

Boas, Hans. 2008b. Resolving form–meaning discrepancies in construction grammar. In Jaakko Leino (ed.), *Constructional reorganization*, 11–36. Amsterdam: John Benjamins.

Boas, Hans. 2014. Lexical and phrasal approaches to argument structure: Two sides of the same coin. *Theoretical Linguistics* 40(1–2). 89–112.

Bouchard, Denis. 1995. *The semantics of syntax: A minimalist approach to grammar*. Chicago: University of Chicago Press.

Bresnan, Joan. 2001. *Lexical-functional syntax*. Oxford: Blackwell.

Bresnan, Joan, Anna Cueni, Tatiana Nikitina & Rolf Harald Baayen. 2007. Predicting the dative alternation. In Gerolf Bouma, Irene Krämer & Joost Zwarts (eds.), *Cognitive foundations of interpretation*, 69–94. Amsterdam: Royal Netherlands Academy of Science.

Bresnan, Joan & Jonni Kanerva. 1989. Locative inversion in Chichewa: A case study of factorization in grammar. *Linguistic Inquiry* 20(1). 1–50.

Buis, Maarten L. 2007. Predict and adjust with logistic regression. *Stata Journal* 7(2). 221–226.

Bybee, Joan. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.

Chang, Winston. 2014. Extrafont: Tools for using fonts.

Cheung, Candice & Richard Larson. 2015. Psych verbs in English and Mandarin. *Natural Language and Linguistic Theory* 33(1). 127–189.

Chung, Taegoo. 1999. External argument and English psychological verbs. *Studies in Generative Grammar* 9(2). 361–380.

Colleman, Timothy. 2006. *De Nederlandse datiefalternantie: Een constructioneel en corpusgebaseerd onderzoek*. Ghent: Ghent University PhD dissertation.

Colleman, Timothy. 2009. Verb disposition in argument structure alternations: A corpus study of the dative alternation in Dutch. *Language Sciences* 31(5). 593–611.

Colleman, Tomothy. 2010. Lectal variation in Constructional Semantics: Benefactive ditransitives in Dutch. In Dirk Geeraerts, Gitte Kristiansen & Yves Peirsman (eds.), *Advances in cognitive sociolinguistics*, 191–221. Berlin: De Gruyter Mouton.

Croft, William. 1986. Surface subject choice of mental verbs. Paper presented at the *Annual Meeting of the Linguistic Society of America*. New York, 27–30 December.

Croft, William. 1991. *Syntactic categories and grammatical relations : The cognitive organization of information*. Chicago: University of Chicago Press.

Croft, William. 1993. Case marking and the semantics of mental verbs. In James Pustejovski (ed.), *Semantics and the lexicon*, 55–72. Dordrecht: Kluwer.

Croft, William. 1998. Event structure in argument linking. In Miriam Butt & Wilhelm Geuder (eds.), *The projection of arguments: Lexical and compositional Factors*, 21–63. Stanford: CSLI Publications.

Croft, William. 2000. *Explaining language change: An evolutionary approach*. Essex: Pearson Education Limited.

Croft, William. 2003. Lexical rules vs. constructions: A false dichotomy. In Hubert Cuyckens, Thomas Berg, René Dirven & Klaus-Uwe Panther (eds.), *Motivation in language : Studies in honor of Günter Radden*, 49–68. Stanford: CSLI Publications.

Croft, William. 2012. *Verbs: Aspect and argument structure*. Oxford: Oxford University Press.

de la Cruz, Juan. 1994. Psych-verbs in old English: From their origin in the lexicon to final syntactic structure. *Studia Anglica Posnaniensia* 28. 37–48.

De Smet, Hendrik & Freek Van de Velde. 2013. Serving two masters: Form-function friction in syntactic amalgams. *Studies in Language* 37(3). 534–565.

de Vries, Matthias & Lamert Allard te Winkel. 1998. *Woordenboek der Nederlandsche taal*.'s-Gravenhage: Nijhoff.

den Besten, Hans. 1982. Some remarks on the ergative hypothesis. *Groninger Arbeiten zur Germanistischen Linguistik* 21. 61–82.

den Boon, Ton & Dirk Geeraerts. 2005. *Van Dale Groot woordenboek van de Nederlandse taal*. 14th edn. Antwerpen/Utrecht: Van Dale Lexicography.

Diessel, Holger. 2015. Usage-based construction grammar. In Ewa Dąbrowska & Dagmar Divjak (eds.), *Handbook of cognitive linguistics*, 296–321. Berlin: De Gruyter Mouton.

Dowty, David. 1991. Thematic proto-roles and argument selection. *Language* 67(3). 547–619.

Filip, Hana. 1996. Psychological predicates and the syntax-semantics interface. In Adele Goldberg (ed.), *Conceptual structure, discourse and language*, 131–147. Stanford: CSLI Publications.

Fox, John. 2003. Effect displays in R for generalised linear models. *Journal of Statistical Software* 8. 1–27.

Geeraerts, Dirk, Gitte Kristiansen & Yves Peirsman (eds.). 2010. *Advances in cognitive sociolinguistics* (Cognitive Linguistics Research [CLR]). Berlin: De Gruyter Mouton.

Goldberg, Adele Eva. 1992. The inherent semantics of argument structure: The case of the English ditransitive. *Cognitive Linguistics* 3. 37–74.

Goldberg, Adele Eva. 1995. *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Goldberg, Adele Eva. 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.

Grafmiller, Jason. 2013. *The semantics of syntactic choice: An analysis of English emotion verbs*. Stanford, CA: Stanford University PhD dissertation.

Gries, Stefan Thomas. 2013. *Statistics for Linguistics with R*. Berlin: De Gruyter Mouton.

Gries, Stefan Thomas. 2015. The most under-used statistical method in corpus linguistics: Multi-level (and mixed-effects) models. *Corpora* 10(1). 95–125.

Gries, Stefan Thomas & Anatol Stefanowitsch. 2004. Extending collostructional analysis: A corpus-based perspective on "alternations." *International journal of corpus linguistics* 9(1). 97–130.

Grimshaw, Jane. 1990. *Argument structure*. Cambridge, MA: MIT press.

Grondelaers, Stefan, Katrien Deygers, Hilde Van Aken, Vicky Van den Heede & Dirk Speelman. 2000. Het CONDIV-corpus geschreven Nederlands [The CONDIV-corpus of written Dutch]. *Nederlandse Taalkunde* 5(4). 356–363.

Haeseryn, Walter, Kirsten Romijn, Guido Geerts, Jaap De Rooij & Maarten van den Toorn. 1997. *Algemene Nederlandse Spraakkunst* [General Dutch Grammar]. Groningen: Nijhoff.

Harrell, Frank. 2013. rms: Regression Modeling Strategies. R package version 4.0-0.

Herschensohn, Julia. 1992. Case marking and French psych-verbs. *Lingvisticæ Investigationes* 16(1). 21–40.

Hoekstra, Eric. 1987. Korte binding en polariteit bij psych-werkwoorden en ditransitieven [Short binding and polarity with psych-verbs and ditransitives]. *Tabu: bulletin voor taalwetenschap* 17. 85–89.

Hopper, Paul & Sandra Annear Thompson. 1980. Transitivity in grammar and discourse. *Language* 56(2). 251–299.

Hothorn, Torsten, Peter Buhlmann, Sandrine Dudoit, Annette Molinaro & Mark Van Der Laan. 2006. Survival ensembles. *Biostatistics* 7(3). 355–373.

Iwata, Seizi. 1995. The distinctive character of psych-verbs as causatives. *Linguistic Analysis* 25. 95–120.

Jackendoff, Ray. 1990. *Semantic structures*. Cambridge, MA: MIT Press.

Kang, Young-Se. 1986. *Korean syntax and Universal Grammar*. Cambridge, MA: Harvard University PhD dissertation.

Kay, Paul. 2005. Argument structure constructions and the argument-adjunct distinction. In Mirjam Fried & Hans Boas (eds.), *Grammatical constructions: Back to the roots*, 71–98. Amsterdam: John Benjamins.

Kay, Paul & Laura Michaelis. 2012. Constructional meaning and compositionality. In Claudia Maienborn, Klaus von Heusinger & Paul Portner (eds.), *Semantics: An international handbook of natural language meaning*, 2271–2296. Berlin: De Gruyter Mouton.

Kelling, Carmen. 2003. French psych verbs and derived nouns. In Miriam Butt & Tracy Holloway King (eds.), *Nominals: Inside and out*, 151–179. Stanford: CSLI Publications.

Klein, Katarina & Silvia Kutscher. 2005. *Lexical economy and case selection of psych-verbs in German. Arbeitspapiere des SFB 282, Theories des Lexicons* [Working papers of the SFB 282, Theory of the Lexico *n*]. *Nr. 122*. Düsseldorf: University of Düsseldorf, Seminar für Allgemeine Sprachwissenschaft. https://ids-pub.bsz-bw.de/frontdoor/index/index/docId/3417 (accessed 3 February 2017).

Klooster, Wim. 2001. *Grammatica van het hedendaags Nederlands: Een volledig overzicht* [Grammar of contemporary Dutch: A complete overview]. Den Haag: Sdu.

Kordini, Valia. 2001. *Psych verb constructions in Modern Greek: A semantic analysis in the hierarchical lexicon*. Essex: Univerisity of Essex PhD dissertation.

Kuhn, Max. 2008. Building predictive models in R using the caret package. *Journal of Statistical Software* 28(5). 1–26.

Lakoff, George. 1977. Linguistic gestalts. *Chicago Linguistic Society* 13. 236–287.

Langacker, Ronald W. 1991. *Foundations of cognitive grammar. Vol.2: Descriptive application*. Stanford: Stanford University Press.

Lasch, Agathe, Conrad Borchling, Gerhard Cordes & Dieter Möhn. 1956. *Mittelniederdeutsches Handwörterbuch*. Neumünster: Wachholtz.

Levin, Beth & Jason Grafmiller. 2012. Do you always fear what frightens you? In Tracy Holloway King & Valeria de Paiva (eds.), *From quirky case to representing space: Papers in Honor of Annie Zaenen*, 21–32. Stanford: CSLI Publications.

Levin, Beth & Malka Rappaport Hovav. 2005. *Argument realization*. Cambridge: Cambridge university press.

Müller, Stefan. 2006. Phrasal or lexical constructions? *Language* 82(4). 850–883.

Müller, Stefan & Stephen Wechsler. 2014. Lexical approaches to argument structure. *Theoretical Linguistics* 40(1–2). 1–76.

Oostdijk, Nelleke, Wim Goedertier, Frank Van Eynde, Louis Boves, Jean-Pierre Martens, Michael Moortgat & Harald Baayen. 2002. Experiences from the Spoken Dutch corpus project. *Proceedings of the third international conference on language resources and evaluation (LREC)*. 340–347. http://www.lrec-conf.org/proceedings/lrec2002/ (accessed 3 February 2017).

Paul, Hermann. 1898. *Prinzipien der Sprachgeschichte*. 3rd edn. Halle an der Saale: Niemeyer.

Pesetsky, David. 1995. *Zero syntax*. Cambridge, MA: The MIT Press.

Philippa, Marlies, Frans Debrabandere, Arend Quak, Tanneke Schoonheim & Nicoline van der Sijs. 2009. *Etymologisch woordenboek van het Nederlands*. Amsterdam: Amsterdam University Press.

Pijnenburg, Wilhelmus. 2001. *Vroegmiddelnederlands woordenboek: Woordenboek van het Nederlands van de dertiende eeuw in hoofdzaak op basis van het Corpus-Gysseling* [Early Middle Dutch dictionary: Dictionary of 13th century Dutch mainly based on the Gysseling Corpus]. Leiden: Instituut voor Nederlandse Lexicologie.

Pijnenburg, Wilhelmus & Tanneke Schoonheim. 2009. *Oudnederlands Woordenboek* [Old Dutch Dictionary]. Leiden: Instituut voor Nederlandse lexicografie.

Pijpops, Dirk & Freek Van de Velde. 2014. A multivariate analysis of the partitive genitive in Dutch Bringing quantitative data into a theoretical discussion. *Corpus Linguistics and Linguistic Theory*. Published online, ahead of print.

Plevoets, Koen. 2008. *Tussen spreek- en standaardtaal: Een corpusgebaseerd onderzoek naar de situationele, regionale en sociale verspreiding van enkele morfosyntactische verschijnselen uit het gesproken Belgisch-Nederlands* [Between spoken and standard language: A corpus-based investigation to the situational, regional and social spread of some morphosyntactic features of spoken Belgian Dutch]. Dissertation University of Leuven.

Postal, Paul Martin. 1970. On the Surface Verb "Remind". *Linguistic Inquiry* 1. 37–120.

Primus, Beatrice. 2002. Proto-roles and case selection in optimality theory. *Arbeiten des Sonderforschungsbereichs 282, Theory des Lexicons, 122*.

Primus, Beatrice. 2004. Protorollen und Verbtyp: Kasusvariaton bei psychischen Verben [Proto-roles and verb type: Case variation with the psychological verbs]. In Rolf Kailuweit & Martin Hummel (eds.), *Semantische Rollen*, 377–401. Tübingen: Gunter Narr Verlag.

Primus, Beatrice. 2006. Mismatches in semantic-role hierarchies and the dimensions of role semantics. In Ina Bornkessel, Matthias Schlesewsky, Bernard Comrie & Angela Friederici (eds.), *Semantic role universals and argument linking*, 53–88. Berlin: Mouton de Gruyter.

R Core Team. 2014. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Rappaport Hovav, Malka & Beth Levin. 2002. Change of state verbs: Implications for theories of argument projection. *Berkeley Linguistics Society* 28. 269–280.

Rogers, Andy. 1974. A transderivational constraint on Richard? In Michael W. La Galey, Robert A. Fox & Anthony Bruck (eds.), *Chicago Linguistic Society* 10. 551–558.

Rohdenburg, Günter. 2003. Cognitive complexity and horror aequi as factors determining the use of interrogative clause linkers in English. In Günter Rohdenburg & Britta Mondorf (eds.), *Determinants of grammatical variation in English*, 205–249. Berlin: Mouton de Gruyter.

Saravanja, Lidija. 2011. Argumentna struktura psiholoških glagola u hrvatskom jeziku [The argument structure of psychological verbs in Croatian]. *Suvremena Lingvistika* 37(2). 241–257.

Speelman, Dirk. 2014. Logistic regression: A confirmatory technique for comparisons in corpus linguistics. In Dylan Glynn & Justyna A. Robinson (eds.), *Corpus methods for semantics: Quantitative studies in polysemy and synonymy*, 487–533. Amsterdam: John Benjamins.

Speelman, Dirk & Dirk Geeraerts. 2009. Causes for causatives: The case of Dutch "doen" and "laten." In Ted Sanders & Eve Sweetser (eds.), *Causal categories in discourse and cognition*, 173–204. Berlin: Mouton de Gruyter.

Stallings, Lynne, Maryellen MacDonald & Padraig O'Seaghdha. 1998. Phrasal ordering constraints in sentence production: Phrase length and verb disposition in heavy-NP shift. *Journal of Memory and Language* 39. 392–417.

Strobl, Carolin, Anne-Laure Boulesteix, Thomas Kneib, Thomas Augustin & Achim Zeileis. 2008. Conditional variable importance for random forests. *BMC Bioinformatics 9* (307) http://www.biomedcentral.com/1471-2105/9/307 (accessed 3 February 2017).

Strobl, Carolin, Anne-Laure Boulesteix, Achim Zeileis & Torsten Hothorn. 2007. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC*

Bioinformatics 8(25). http://www.biomedcentral.com/1471-2105/8/25 (accessed 2 February 2017).

Szmrecsanyi, Benedikt, Jason Grafmiller, Benedikt Heller & Melanie Röthlisberger. 2016. Around the world in three alternations: Modeling syntactic variation in varieties of English. *English World-Wide* 37(2). 109–137.

Theijssen, Daphne. 2012. *Making choices: Modelling the English dative alternation*. Nijmegen: Radboud University PhD dissertation.

Van de Velde, Freek. 2004. De Middelnederlandse onpersoonlijke constructie en haar grammaticale concurrenten: Semantische motivering van de argumentstructuur [The Middle Dutch impersonal construction and her grammatical competitors: Semantic motivation of the argument structure]. *Nederlandse Taalkunde* 9(1). 48–76.

Van de Velde, Freek. 2014. Degeneracy: The maintenance of constructional networks. In Ronny Boogaart, Timothy Colleman & Gijsbert Rutten (eds.), *Extending the Scope of Construction Grammar*, Vol. 1, 141–179. Berlin: Mouton de Gruyter.

van der Horst, Joop. 1985. Verkenning van de onpersoonlijke constructies [Exploration of the impersonal constructions]. *Tijdschrift voor Nederlandse Taal- en Letterkunde* 101. 35–63.

van der Horst, Joop. 2008. *Geschiedenis van de Nederlandse syntaxis* [History of Dutch syntax]. Leuven: Universitaire Pers Leuven.

van Eerten, Laura. 2007. Over het Corpus Gesproken Nederlands [On the Corpus of Spoken Dutch]. *Nederlandse Taalkunde* 12(3). 194–215.

van Trijp, Remi. 2015. Cognitive vs. generative construction grammar. The case of coercion and argument structure. *Cognitive Linguistics* 26(4). 613–632.

Van Valin, Robert Junior. 2004. Semantic macroroles in role and reference Grammar. In Rolf Kailuweit & Martin Hummel (eds.), *Semantische Rollen*, 62–82. Tübingen: Gunter Narr.

Vanhoe, Henk. 2002. Aspects of the syntax of psychological verbs in Spanish: A lexical functional analysis. In Miriam Butt & Tracy H. King (eds.), *Proceedings of the LFG02 Conference*, 373–389. Stanford: CSLI Publications.

Venables, William & Brian Ripley. 2002. *Modern applied statistics with S*. 4th edn. New York: Springer.

Verhoeven, Elisabeth. 2010. Agentivity and stativity in experiencer verbs: Implications for a typology of verb classes. *Linguistic Typology* 14(2–3). 213–251.

Verwijs, Eelco & Jacob Verdam. 1991. *Middelnederlandsch woordenboek*. Zedelgem: Zedelgem Flandria Nostra.

Wasow, Thomas. 2002. *Postverbal behavior*. Stanford: CSLI Publications.

Willems, Annelore & Gert De Sutter. 2015. Reassessing the effect of the complexity principle on PP Placement in Dutch. *Nederlandse Taalkunde* 20(3). 339–366.

Willems, Annelore & Gert De Sutter. 2016. Understanding PP placement in written Dutch: A corpus-based multifactorial investigation of the principal syntactic, semantic and discursive determinants. In Johannes Wahle, Marisa Köllner, Harald Baayen, Gerhard Jäger & Tineke Baayen-Oudshoorn (eds.), *Proceedings of the 6th conference on quantitative investigations in theoretical linguistics*. Tübingen. https://publikationen.uni-tuebingen.de/xmlui/handle/10900/64864 (accessed 3 February 2017).

Yang, Byong-Seon. 1994. *Morphosyntactic phenomena of Korean in Role and Reference Grammar: Psych-verb construction, inflectional verb morphemes, complex sentences, and relative clauses*. Buffalo, NY: State University of New York PhD dissertation.

Zaenen, Annie. 1993. Unaccusativity in Dutch: Integrating syntax and lexical semantics. In James Pustejovsky (ed.), *Semantics and the lexicon*, 129–161. Dordrecht: Kluwer.

Zifonun, Gisela, Ludger Hoffmann & Bruno Strecker. 1997. *Grammatik der deutschen Sprache* [Grammar of the German language]. Berlin: Mouton de Gruyter.

Zwart, Jan-Wouter. 2011. *The syntax of Dutch*. Cambridge: Cambridge University Press.

Appendix Examples in Table 2

(16) *Wij storen niemand en   zelf worden we ook niet gestoord.*
     we disturb no-one   and self are   we also not disturbed
     'We do not disturb anyone and we are not disturbed ourselves.'

(17) *Tony Herreman interesseert ons wel.*
     Tony Herreman interests   us   well
     'Tony Herreman does interest us.'

(18) *De Arabische wereld ergerde zich indertijd   dood aan de grote*
     the Arabic   world annoyed itself at.the.time dead to the large
     Israeli   delegatie.
     Israelische delegatie
     'The Arabic world was, at the time, extremely annoyed by the large Israeli delegation.'

(19) *Zowel de gemeente   Maasbracht als de ondernemers [...] ergeren*
     Both the municipality Maasbracht as the entrepreneurs   annoyed
     zich   over het wrak   [...]
     themselves about the wreckage
     'Both the municipality of Maasbracht and the entrepreneurs [...] were annoyed by the wreckage [...].'

(20) *Auto's interesseren me eigenlijk niet zo   veel.*
     Cars interest   me actually not that much
     'Actually, car's don't really interest me that much.'

(21) *Wat ons verbaasde was de hoge kostprijs van het apparaat,*
     What us amazed   was the high cost   of the machine
     rond   het half miljoen.
     around the   half million
     'What amazed us, was the high cost of the machine, about half a million.'

(22) *[...] wanneer hij zich   ergert   aan het breken   met de regel [...]*

    when   he himself annoys to   the breaking with the rule
'When he is annoyed by the breaking of the rule,...'

(23) *Je   houdt namelijk weinig vrienden over als ned.-belgië*
You hold   namely few   friends   left   if   the Netherlands-Belgium

je   geen bal interesseert.
you no   ball   interests
'You see, you won't be left with many friends if the soccer game the
Netherlands vs. Belgium doesn't interest you at all.'

(24) *[...] omdat hij zich   ergerde aan het gegroet.*
      because he himself annoyed to   the greeting
'Because all the greeting was annoying him.'

(25) *De landelijke politiek interesseert me geen klap.*
The national   politics interests   me no   smack
'National politics do not interest me at all.'

(26) *Eén dag vergaderen met de raad van bestuur* -   *je ergert*
one day meeting   with the board of   management you annoy
je   gek aan de inefficiëntie.
yourself crazy to   the inefficiency
'Have a meeting of one day with the executive board, and you'll be driven
crazy by the inefficiency.'

(27) *Die hypocrisie stoort me enorm*,   *want   ik schrijf slechts*
That hypocrisy disturbs me immensely because I   write   only
over de vragen   die   kinderen stellen.
about the questions those children pose
'That hypocrisy bothers me immensely, because I only write about the
questions that children ask.'

(28) *Dat jou iets   ergert ... is weer een ander verhaal.*
that jou something annoys   is again a   other   story
'That something annoys you, is another story altogether.'

(29) *Ik ergerde er   mij   een beetje aan dit soort dingen.*
I   annoyed there myself a   bit   to   that kind things
'That kind of stuff annoyed me a bit'.

(30) *En　op het internet vond ik vrienden met wie　ik kon　praten*
And on the internet found I friends　with whom I　could talk

over　wat　mij interesseerde.
about what me interested
'And on the internet, I found friends with whom I could talk about the things that interested me.'

(31) *Taihuttu ergert zich　eraan　dat de afgelaste wedstrijden*
Taihuttu annoys himself thereto　that the cancelled games
zo　laat　worden ingehaald.
so　late　are　rescheduled
'Taihuttu is annoyed that the canceled games are rescheduled so late.'

(32) *Dat　de　inhoud ervan bij "buitenstaanders" bekend was,*
that　the　contents thereof with outsiders　known were
ergerde hem.
annoyed him

'That its contents were known to "outsiders" annoyed him.'

(33) *Al　besef ik dat ik daar moeilijkheden mee krijg, maar dat*
Although realize I that I there difficulties　with get　but　that
interesseert me niet.
interests　me not
'Although I realize that that will get me into trouble, but I don't care about that.'

(34) *Wat　me nu　vooral　stoort,　is het feit dat we twee goals*
What me now especially disturbs is the fact that we two　goals
weggeven uit　eigen fout.
away-give from own　fault
'What is especially bothering me now, is the fact that we give two goals away, which were our own fault.'

Variance Inflation Factors (VIF)

Table 6: Variance Inflation Factors of the regression model in Table 4 (r.l. = reference level).

Explanatory variable	level	VIF	Explanatory variable	level	VIF
Verb	interesseren	r.l.	Tense	present	r.l.
	ergeren	1.88		past	1.12
	storen	1.86		future	1.08
	verbazen	2.25		conditional	1.05
Stimulus-Topicality	third person pronoun	r.l.	Negation	without	r.l.
	first-person	1.33		with	1.11
	second-person	1.16	Stimulus-Number	singular	r.l.
	definite noun	2.13		plural	1.33
	indefinite noun	1.62	Register	e-mail	r.l.
	subordinate clause	2.27		chat	1.32
Experiencer-Topicality	first person	r.l.		informal speech	1.29
	second person	1.24		formal speech	1.37
	third person pronoun	1.37		formal speech	1.37
	definite noun	1.42		mass newspaper	1.75
	indefinite noun	1.16		quality newspaper	1.38
Country	Belgium	r.l.			
	the Netherlands	1.19			
Stimulus-Animacy	proposition	r.l.			
	animate	2.58			
	inanimate	1.89			
	concrete	2.67			
	event	1.73			
	abstract	2.25			